THE UNIVERSITY *of* EDINBURGH

The role of C/EBPβ in two luminal progenitor populations in the mouse mammary gland

Agnes Zay

Thesis presented for the degree of Doctor of Philosophy
The University of Edinburgh
2012

Declaration

I hereby declare that this thesis is of my own composition, and that it contains no material previously submitted for the award of another degree. The work presented in this thesis is my own, unless stated otherwise.

Agnes Zay

Acknowledgements

I would like to express my deepest gratitude to my supervisor, Prof. Claus Nerlov, for his mentorship and support over the past four years. He created an excellent research and working environment that fostered independent inquiry and personal growth.

I would like to thank the members of my PhD supervisory committee, Dr. Val Wilson and Prof. Alexander Medvinsky, for their support and helpful comments.

I would also like to thank Dr. Susana Garcia Silva, who took me under her wing and taught me about mammary gland biology. Her pioneering work on the mammary gland led to the development of the projects presented in this thesis.

As well, I would like to acknowledge my fellow lab members for their help and support: Dr. Mario Buono, Dr. Amit Grover, Dr. Alejandra Sanjuan-Pla, Dr. Roy Drissen, Dr. Adriana Gambardella, Chinnavuth Vatanashevanopakorn, Martina Pigoni, Julie Scotton and Susan Moore.

I would like to thank Dr. Janus Schou Jakobsen and Dr. Pablo Navarro for helpful discussions on the development of the ChIP method, and Dr. Simon Tomlinson and Florian Halbritter for their assistance with the bioinformatics analysis.

I would also like to thank the staff of the SCRM animal unit who took such excellent care of my mice: Jamie Todd, David McIver, Laura Walker and Carol Manson.

I would also like to thank the flow labs (Simon Monard, Dr. Olivia Rodrigues, Fiona Rossi, Dr. Martin Waterfall) for helping me through the numerous hours of sorting needed to generate material for the ChIP experiments. I would also like to thank Dr. Valeria Berno for her expertise and assistance with imaging/microscopy. As well, I am grateful to Jon Rans, Teresa O'Connor, Helen Henderson and Marilyn Thomson for running an excellent tissue culture facility; Robbie McCaig for managing all of our ordering; Audrey Wright and Euan McPherson for stocking our cupboards with sterile plastics and glassware; and Pat Hope, Vaila O'Connor, Dr. Caroline Proctor and Kelly Douglas for administrative support.

Pentultimately, I would like to thank my Family. It has been difficult to live so far away from you these past four years, but your unconditional love and support kept me focused and committed to this project. I love you all so very much.

And most importantly, I would like to thank the love of my life, my Babs, Javier, for our life together. Words cannot describe the love and respect I have for you, and I thank you for your unconditional love, unwavering support, endless patience, and brilliant scientific mind.


In dedication to
my family
You are deeply
loved and cherished

Abstract

The mammary gland is a branched epithelial organ comprised of myoepithelial, ductal and alveolar cells that are derived from resident stem and progenitor cells. The progression from mammary gland stem cell(s) to the differentiated mammary gland cell types is poorly understood. Here, I describe the identification and characterization of two luminal progenitor cell populations in the mouse mammary gland, and investigate the role of the transcription factor C/EBPβ in their development. In Chapter 2, I describe the isolation of two luminal progenitor cell populations (Sca1+ and Sca1- luminal cells) and show that they are differentially primed in their gene expression towards ductal and alveolar cell fates, respectively. Furthermore, I show that *in vivo* genetic priming affects the *in vitro* differentiation potential of Sca1+ and Sca1- luminal cells. In Chapter 3, I show that C/EBPβ is required for the appropriate specification of ductal and alveolar lineages, and in its absence, alveolar lineage priming is lost, and ductal lineage priming is up-regulated in both Sca1+ and Sca1- cells. Preliminary data also shows that in addition to severe proliferation defects, the changes in *in vivo* lineage priming in *Cebpb*-/- mice also affect the *in vitro* differentiation potential of *Cebpb*-/- Sca1+ and Sca1- luminal progenitors. Lastly, in Chapter 4, I describe the genome-wide binding characteristics of C/EBPβ in Sca1+, Sca1- and P16.5 alveolar cells. These experiments reveal that genome-wide C/EBPβ occupancy is correlated with alveolar cells fate, and that C/EBPβ target genes perform distinct cellular functions in alveolar cells (Sca1- cells and P16.5). Furthermore, I show that *Elf5* is directly regulated by C/EBPβ, and posit that direct regulation of *Elf5* by C/EBPβ may be one mechanism through which C/EBPβ exerts its alveolar cell fate programming.

Table of contents

Figures and Tables

Figures

Tables

Abbreviations

7-AAD	7-aminoactinomycinD
A-genes	Alveolar genes
AD	Activation domain
ALP	Alveolar luminal progenitor cell
AREG	Amphiregulin
ATF	Activating transcription factor
ATP	Adenosine triphosphate
ATP6V1B1	V-type proton ATPase subunit B
bHLH	Basic helix-loop-helix
BMP	Bone morphogenic protein
BMP1RA	BMP receptor 1A
BP	Base pair
BrdU	Bromo deoxyuridine
BSA	Bovine serum albumin
bZIP	Basic leucine zipper
C	Control ChIP (IgG)
C/EBP	CCAAT enhancer binding protein
CaMK	Calcium calmodulin-dependent kinase
cAMP	Cyclic adenosine monophosphate
CASP1	Interleukin-1β converting enzyme
CBP	CREB binding protein
CCND1	Cyclin D1
CD	Cluster of differentiation
CD133	Prominin 1
CD29	Beta 1 integrin
CD49b	Alpha 2 integrin
CD49f	Alpha 6 integrin
CD61	Beta 3 intergin
CFC	Colony forming cells
CFP	Cyan fluorescent protein
ChIP	Chromatin immunoprecipitation
ChIP-Seq	Chromatin immunoprecipitation sequencing
Cp	Crossing point
CREB	cAMP response element
CSN2	Beta casein
CTGF	Connective tissue growth factor
CUG-BP1	CUG binding protein 1

Cx26	Connexin 26
D-genes	Ductal genes
DBD	DNA binding domain
DKK1	Dickkopf
DLR	Dual luciferase reporter assay
DLP	Ductal luminal progenitor cell
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
DR	Death receptor
E	Embryonic day
EDTA	Ethylenediaminetetraacetic acid
EGF	Epidermal growth factor
EGFR	EGF receptor
ELF5	E74-like factor 5
ES cells	Embryonic stem cells
ESR	Oestrogen receptor
ESR1	Oestrogen receptor alpha
ESR2	Oestrogen receptor beta
FACS	Fluorescence-activated cell sorting
FCS	Fetal calf serum
FDR	False detection rate
FGF	Fibroblast growth factor
FMO	Fluoresence minus one control
FOXA1	Forkhead box protein A1
FSC	Forward scatter
FSC-A	Forward scatter area
FSC-H	Forward scatter height
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
gDNA	Genomic deoxynucleic acid
GFP	Green fluorescent protein
GH	Growth hormone
GHR	Growth hormone receptor
GO	Gene ontology
GR	Glucocorticoid receptor
HB2	Ductal carcinoma in situ cell line
HBB-B1	β-globin
HBSS	Hank's balanced salt solution
HC11	Mammary epithelial cell line

HER4	ERBB4 receptor tyrosine kinase
HH	Hedgehog
HI	High
HIF1A	Hypoxia inducible factor 1A
HSC	Hematopoietic stem cell
IGF1	Insulin-like growth factor 1
IKK	Inhibitor of kappa beta kinase
IKKDR	Inhibitor of kappa beta kinase death receptor
IL	Interleukin
IMDM	Iscove's Modified Dulbecco's Medium
IP	Target ChIP (e.g., C/EBPβ)
K	Keratin
KB	Kilobase
kDa	Kilodalton
KLF4	Krüppel-like factor 4
L-genes	Luminal genes
LAP	Liver enriched transcriptional activator protein
LB	Limb bud
LH	Luteinizing hormone
LIF	Leukemia inhibitory factor
LIN	Lineage
LIP	Liver enriched transcriptional inhibitor protein
LO	Low
LTF	Lactoferrin/lactotransferrin
M	Marker
Ma-CFC	Mammary colony forming cell
MACS	Model-bsed analysis of ChIP-Seq
MAPK	Mitogen activated protein kinase
MAX	Maximum
MB	Mammary bud
MCSF1R	Marcophage colony stimulating factor 1 receptor
MEC	Mammary epithelial cell
MET	Hepatocyte growth factor receptor
MG-SP	Mammary gland side population
Mk/E	Megakaryocytic/Erythroid
ML	Milk line
MMP	Matrix metalloproteinase
MMTV	Mouse mammary tumour virus
mRNA	Messenger RNA
MRU	Mammary repopulating unit
MSX2	Muscle segmentation homeobox 2

MYO	Myoepithelial cells
N	Number
NEG	Negative
N.D.	Not determined
N.S.	Not significant (as determined by Student's T-Test)
NF	Nuclear factor
NP40	Nonyl phenoxypolyethoxylethanol
NRG3	Neuroregulin 3
P	Placode
P/S	Penicillin/streptomycin
P16.5	Mammary luminal cells at day 16.5 or pregnancy
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PGR	Progesterone receptor gene
PH domain	Plekstrin homology domain
PI3K	Phosphatidylinositol 3 kinase
PIC	Protease inhibitor cocktail
PKA	Protein kinase A
PKB	Protein kinase B
PKC	Protein kinase C
PR	Progesterone receptor
PRA	Progesterone receptor alpha
PRB	Progesterone receptor beta
PRL	Prolactin
PRLR	Prolactin receptor
PSR	Phophatidylserine receptor
PTHLH	Parathyroid hormone-related peptide
qPCR	Quantitative polymerase chain reaction
qRT-PCR	Quantitative reverse transcriptase polymerase chain reaction
RANKL	Receptor activator of NF κβ ligand
RD	Repressor domain
rhEGF	Recombinant human epidermal growth factor
rhbFGF	Recombinant
RMA	Robust multiarray analysis
RNA	Ribonucleic acid
RPM	Rotation per minute
RSK2	Ribosomal S6 kinase 2
SA	Streptavidin
Sca1	Stem cell antigen 1
SCRM	Scottish Centre for Regenerative Medicine

SD	Standard deviation
SDS	Sodium dodecyl sulfate
siRNA	Small interfering RNA
SMA	Smooth muscle actin
SMO	Smoothened
SREBP1	Sterol regulatory element binding protein 1
SSC	Side scatter
SSC-A	Side scatter area
STAT5	Signal transducer and activator 5
T47D	Human breast cancer cell line
TBX3	T-box transcription factor
TEB	Terminal end bud
TFAS	Transcription factor associated score
TFBS	Transcription factor binding site
TGFβ3	Transforming growth factor beta 3
TIMP	Tissue inhibitor of matrix metalloproteinase
TNC	Tenascin-C
TNF	Tumour necrosis factor
TNFR1	Tumour necrosis factor receptor 1
TRPM2	Testosterone repressed prostate message 2
TSS	Transcription start site
UMS	Ulnar mammary syndrome
UPL	Universal probe library
v/v	Volume to volume
w/v	Weight to volume
WAP	Whey acidic protein
WNT	Wingless-related MMTV integration site
WT	Wild type
XDH	Xanthine dehydrogenase
YFP	Yellow fluorescent protein

1 Introduction

1.1 The mouse mammary gland

1.1.1 Embryonic development

Overview

The mouse mammary gland begins to form during embryonic development from a small number of ectoderm-derived epithelial cells (Sakakura, 1987). Between embryonic day 10 (E10) and E11, these cells form two tracks, often referred to as the milk or mammary line, that extend from the thoracic to the inguinal region along the ventral body surface. Over the next 24 hours, five pairs of lens-shaped epithelial thickenings, known as placodes, form along the mammary line at the future sites of the glands (Veltmaat et al., 2003; Robinson, 2007). These placodes, attached to the epidermis by a narrow stalk, slowly increase in size between E13 and E15, and invaginate the underlying dermis to form bulb-shaped buds. Between E16 and E18, the buds elongate and form a primitive ductal system by branching into the developing subcutaneous fat pads. These rudimentary structures persist unchanged until puberty. Also around E16, specialized mesenchymal cells condense into concentric layers around the rudimentary mammary glands, and begin to express steroid receptors (Heuberger et al., 1982; Robinson et al., 1999). Concomitantly, the surface epithelia associated with the glands differentiates into specialized nipple skin cells (Foley et al., 2001). Mouse mammary gland development is indistinguishable between males and females before E13. On E13, the male testes begin to secrete androgens, which, over the next two days, cause the mammary epithelium to degenerate and detach from the nipple (Richert et al., 2000). These developmental steps are discussed in more detail below and summarized in Figure 1.1 [adapted from (Robinson, 2007)].

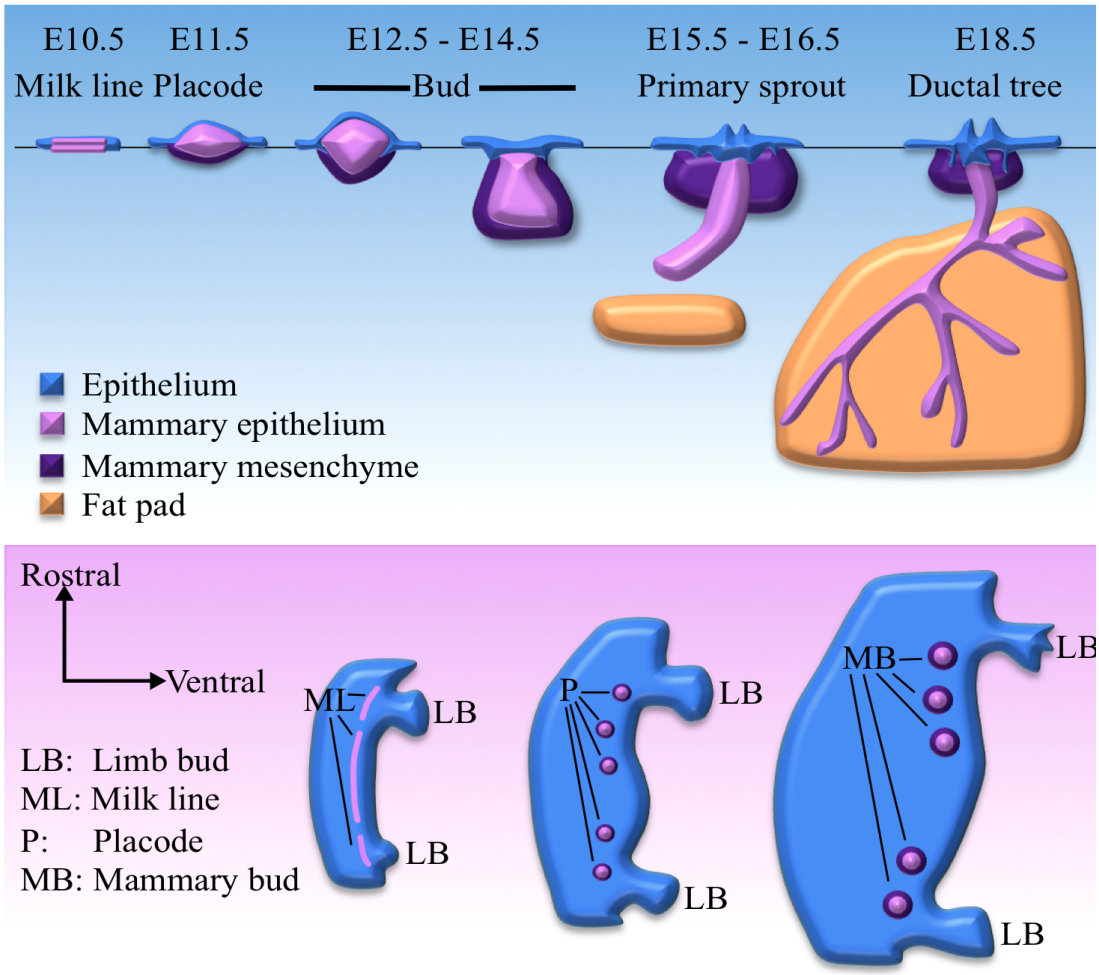


Figure 1.1 Key developmental stages in the embryonic mammary gland. The top schematic diagram depicts the formation of the embryonic mammary ductal tree. The first specification of the mammary gland occurs around embryonic day 10.5 (E10.5) with the formation of two lateral tracks along the ventral body surface, referred to as the milk line. By E11.5, mammary placodes form at precise positions along these tracks, and grow to form mammary buds (E12.5-E14.5). The mammary buds sprout into the underlying dermis between E15.5 and E16.5, and branch into the subcutaneous mammary fat pad by E18.5. The bottom schematic diagram shows the positioning of the mammary buds in the developing mouse embryo. Two milk lines (ML) form on the rostral-caudal axis along the ventral body surface. Five pairs of precisely positioned mammary placodes (P) form along the milk lines, each of which develops into a mammary bud (MB). This figure is adapted from Robinson, 2007.

Initiation of mammopoiesis

The first steps of mammopoiesis in the mouse embryo are initiated by the activation of WNT (wingless-related mouse mammary tumor virus (MMTV) integration site), FGF (fibroblast growth factor) and HH (hedgehog) signalling pathways (Chu et al., 2004; Veltmaat et al., 2004). The earliest marker of the mammary line, appearing at E10.25, is the WNT and FGF target gene T-box transcription factor 3 (*Tbx3*) (Eblaghie et al., 2004). TBX3 was first noted for its involvement in mammary gland development in patients with ulnar mammary syndrome (UMS) (Bamshad et al., 1999). UMS patients suffer from hypoplasia in the mammary glands, as well as in other organs, and mutations in the deoxynucleic acid (DNA) binding domain of the *Tbx3* gene have been associated with this disease (Bamshad et al., 1999). In mouse development, deletion of *Tbx3* leads to complete absence of prenatal mammary epithelium (Davenport et al., 2003). Canonical WNT activity can also be detected in the mesenchymal cells along the mammary line from E10.5, with increased expression in and around the placodes as they form (DasGupta and Fuchs, 1999). In this study, canonical WNT activity was detected using the TOP-Gal reporter mouse, which contains the *lacZ* gene under the control of the *c-Fos* minimal promoter. Upon WNT stimulation, the *lacZ* gene product, beta galactosidase, is expressed and its activity can be detected using an X-gal assay (DasGupta and Fuchs, 1999). In the E11.25 embryo, expression of *Wnt10b* and *Wnt6* can be found along the whole mammary line, while *Wnt10a* is found only in the placodes (Veltmaat et al., 2004). As well, another WNT target gene, *Lef1* is dynamically expressed along the milk line from E10.5 (Mailleux et al., 2002). Over-activation of WNT signalling accelerates the formation of placodes in their normal location, as well as induces the formation of ectopic placode-like structures (Chu et al., 2004), while keratin 14 (K14)-driven expression of the WNT inhibitor Dickkopf (*Dkk1*) leads to complete absence of any mammary gland structures (Andl et al., 2002; Chu et al., 2004). LEF1 and TBX3 have been identified as the most likely effectors of early mammopoiesis due to the absence of placode formation upon deletion of each of these factors (van Genderen et al., 1994; Davenport et al., 2003; Renard et al., 2007). Using *Lef1* as a marker of placode formation also revealed that the mammary placodes form in an

asynchronous way (Mailleux et al., 2002). Between E11 and E11.5, placode 3 begins to form first, followed by placode 4, then 1 and 5, and finally by placode 2.

Mammary placode formation

Unique and complex signalling interactions are involved with the formation of each placode pair (Mailleux et al., 2002; Veltmaat et al., 2006). For example, genetic deletion of *Fgf10* or its receptor *Fgfr2b* prevents the development of placodes 1,2,3 and 5, but not placode 4 (Mailleux et al., 2002). Deletion of *Gli3*, a hedgehog signalling pathway inhibitor, prevents the formation of placodes 3 and 5, while the other placodes develop normally (Hatsell and Cowin, 2006; Veltmaat et al., 2006). On the other hand, deletion of the hedgehog signalling pathway activator Smoothened (*Smo*) leads to ectopic mammary gland formation on the ventral skin of the mouse embryo (Gritli-Linde et al., 2007). In the absence of neuroregulin 3 (*Nrg3*), a ligand for ERBB4 receptor tyrosine kinase (HER4), development of placode 3 is inhibited and ectopic placode formation on the dorsal side of placode 4 occurs (Howard et al., 2005). Another level of complexity is added to the development of the placodes through the regulation of dorso-ventral positioning. For example, precise positioning of placode 3 is initiated on E10.5 by reciprocal expression of *Tbx3/15* and *Bmp4* in the mesenchymal cells surrounding the mammary epithelium (Cho et al., 2006). High *Tbx3/15* and low *Bmp4* expression mark the ventral boundary of placode 3, while high *Bmp4* and low *Tbx3/15* expression mark the dorsal boundary. TBX3, TBX15 and BMP4 each regulate the regional expression of *Lef1* and specify the boundaries of this placode. Although our understanding of the spatial regulation of placode formation is not complete, it is clear that this process requires a complex cooperation and coordination of signals.

Mammary bud formation

Morphologically distinct epithelial buds, formed by E13.5, initiate the reorganization of the proximal mesenchyme into tightly packed, concentric layers of cells around the mammary epithelium. At this time point in development, these specialized cells,

referred to as the primary mammary mesenchyme, begin to express hormone receptors and the matrix protein tenascin C (TNC) (Heuberger et al., 1982; Dunbar et al., 1999; Robinson et al., 1999). The epithelial signal that induces reorganization of the proximal mesenchyme is parathyroid hormone-related peptide (PTHLH) (Dunbar et al., 1999), a peptide hormone more commonly known for its role in calcium homeostasis (Mannstadt et al., 1999). *Pthlh* is expressed from E11.5 in the mammary epithelium, whereas its receptor, *Pthr1* is expressed in all of the epidermis (Dunbar et al., 1999). In the absence of PTHLH or its receptor, the mammary placodes progress into mammary buds, but the buds are unable to elongate and push through the underlying mesenchyme into the subcutaneous fat pads (Wysolmerski et al., 1998). The absence of PTHLH signalling also affects cytokeratin expression in the mammary buds, where it changes from a mammary-specific pattern (Asch and Asch, 1985; Sun et al., 2010) to a skin-cell specific pattern (Foley et al., 2001). As well, the expression of androgen receptor in the primary mammary mesenchyme is lost. On the other hand, ectopic expression of PTHLH induces androgen receptor expression and reorganization of the underlying mesenchyme, and leads to differentiation of the surrounding epidermis into specialized nipple cells (Dunbar et al., 1999; Foley et al., 2001). Taken together, these data show that PTHLH signalling is required for the formation of the primary mammary mesenchyme and for the androgen-mediated growth of the mammary bud.

Embryonic branching morphogenesis

Induction of mammary bud elongation, branch initiation and nipple skin specialization are also dependent on PTHLH signalling (Hens et al., 2007). PTHLH is secreted by the mammary epithelium, and promotes the expression of bone morphogenic protein (BMP) receptor 1A (*Bmpr1a*) in the primary mammary mesenchyme. In turn, BMP4 binds BMPR1A in the mesenchymal cells and initiates branching of the elongated mammary epithelial buds (Hens et al., 2007). This mechanism was elucidated through the observations that addition of BMP4 to mammary bud explant cultures rescues the defect in ductal morphogenesis in PTHLH-deficient mice, and conversely, that addition of Noggin, a BMP antagonist,

reduces ductal branching by 50 % in wild type mammary buds (Hens et al., 2007). Concurrent to branch initiation of the mammary epithelium is the synergistic activation of muscle segmentation homeobox 2 (*Msx2*) by PTHLH and BMP in the adjacent epidermis (Hens et al., 2007). MSX2 inhibits hair follicle formation and induces the specialization of the nipple cells. Taken together, the activation of mesenchymal BMP signalling by the secretion of PTHLH from the mammary epithelium stimulates mammary bud elongation and branching into the underlying mammary fat pads, as well as mediates nipple formation by MSX2-dependent inhibition of hair follicle development. Development of the mammary gland arrests on E18.5, and will not resume until puberty (at ~3 weeks of age in mice).

1.1.2 Postnatal development

Overview

In the first three weeks after parturition, the small ductal trees that developed in the embryo grow allometrically (Watson and Khaled, 2008). From the onset of puberty (from ~3 weeks of age in mouse), the development of the mammary gland is regulated primarily by the sequential actions of growth hormone (GH), oestrogen, progesterone and prolactin (Figure 1.2) (Brisken and O'Malley, 2010). During puberty, the cyclic release of oestrogen and progesterone stimulate the ducts to grow and branch. Most hormone responsive cells are located in specialized structures at the ends of the ducts called terminal end buds (TEBs) (Fendrick et al., 1998; Joshi et al., 2012). Proliferation in the TEBs leads to ductal elongation, while clefting of TEBs generates branches. Over the next few weeks, the virgin mammary gland continues to develop into a regularly branched epithelial tree that completely fills the mammary fat pad, and is ready for the next stage of development, terminal differentiation of the secretory alveolar cells.

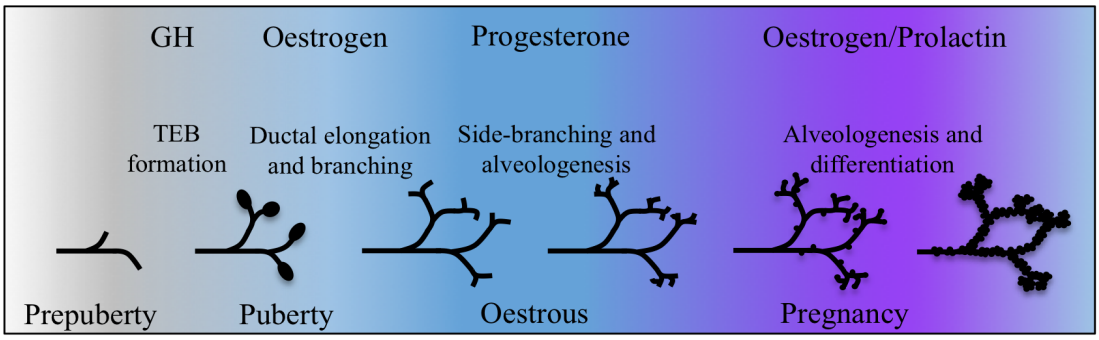


Figure 1.2 Hormone actions during postnatal mammary gland development. The mammary gland becomes hormone responsive at puberty, and its subsequent development is regulated by the sequential actions of growth hormone (GH), oestrogen, progesterone and prolactin. Terminal end bud (TEB) development and ductal morphogenesis are regulated primarily by indirect GH signalling in early postnatal mammary gland development. At puberty, oestrogen signalling promotes further development of the ductal epithelium by regulating ductal elongation and branching. The cyclical release of oestrogen and progesterone in the adult virgin mammary gland leads to the proliferation and apoptosis of a subset of mammary epithelial cells with each oestrous cycle. During early pregnancy, progesterone promotes further ductal side-branching, whereas during mid-late pregnancy, increased levels of oestrogen and prolactin promote alveolar proliferation and differentiation. The cessation of suckling leads to a sharp drop in prolactin levels and remodelling of the mammary gland, known as involution, to a state that resembles its pre-pregnancy state. This figure is based on Brisken and O'Malley, 2010.

Development of the terminal end buds

TEBs form shortly after the onset of puberty, and are comprised of several layers of body cells that are surrounded by an outer layer of cap cells (Sternlicht, 2006). Elongation of the ducts results from proliferation within the TEBs, where cap cells contribute to the growth of the basal myoepithelial cell layer, and body cells contribute to the growth of the luminal cell layer (Hennighausen and Robinson, 2005). A high rate of apoptosis occurs in the body cells of TEBs (~2x that observed in the mature ductal epithelium), which is posited to be the mechanism for lumen formation (Humphreys et al., 1996; Richert et al., 2000; Watson and Khaled, 2008). Clefting in the TEBs leads to bifurcation of the ducts that generate additional branches (Watson and Khaled, 2008). The formation of TEBs at the onset of puberty is essential for the rapid, hormone-dependent growth of the mammary gland. Interactions between the mammary epithelium and the fat pad are important for TEB development (Couldrey et al., 2002). In mice lacking white adipose tissue, ductal development is severely disrupted and TEBs do not form (Couldrey et al., 2002). Interestingly, these mice are still capable of producing milk, suggesting that although adipocytes are necessary for ductal growth and branching, they are not required for the terminal differentiation of alveolar cells. The expression of Rras2 within the mammary epithelium is also necessary for normal TEB formation and ductal development. R-RAS2, a member of the RAS GTPase superfamily, is expressed at low levels in the mammary gland, with increased expression observed during puberty and pregnancy (Larive et al., 2012). Mammary gland development is normal until puberty in *Rras2*-/- mice. However, at puberty, fewer TEBs develop, which leads to a delay in ductal elongation and branching (Larive et al., 2012). Another important factor for TEB formation is GATA3, a zinc-finger transcription factor that is normally expressed in the body cells of TEBs. MMTV-Cre driven excision of *Gata3* leads to defects in ductal elongation and branching during puberty (Asselin-Labat et al., 2006; Kouros-Mehr et al., 2006). GATA3 is also critical for the development of alveolar cells, and its absence leads to a block in the differentiation of alveolar progenitor cells into milk producing cells (Asselin-Labat et al., 2006). Loss of GATA3 is further associated with reduced expression of oestrogen receptor alpha

(ESR1) in the mammary epithelium. Kouros-Mehr proposed that forkhead box protein A1 (FOXA1), a transcription factor recently shown to regulate the DNA binding activity of ESR1 (Lupien et al., 2008; Hurtado et al., 2011), is a direct target gene of GATA3, and thereby may mediate the cross-talk between GATA3 and oestrogen signalling (Kouros-Mehr et al., 2006).

Growth hormone signalling

Growth Hormone (GH) signalling is primarily important for TEB development and ductal morphogenesis. Historically, GH was thought to mimic the lactogenic effects of prolactin due to their structural similarity (Kleinberg et al., 1990). However, Lyons was the first to hypothesize that GH may play an important role during puberty (in TEB formation and branching), whereas prolactin is mostly involved in lactogenesis (Lyons et al., 1958). The first study to demonstrate this compared the effects of oestrogen, oestrogen and GH, and oestrogen and prolactin on the mammary glands of hypophysectomised rats (Kleinberg et al., 1990). These experiments showed that oestrogen and GH, but not oestrogen alone, nor oestrogen and prolactin, can increase pubertal mammary gland development, as evidenced by increased TEB formation and ductal branching. Also, treatment with GH, or its downstream effector insulin-like growth factor 1 (IGF1), led to a 2-fold increase in the size of the mammary epithelium in rhesus monkeys (Murphy et al., 1987; Ng et al., 1997). More recently, through the transplantation of GH receptor (GHR) null mammary epithelia into wild type stroma, it was demonstrated that growth hormone signalling is not directly involved in the development of the mammary epithelium (Gallego et al., 2001). Rather, GH must interact through an intermediate messenger in the stroma to cause the GH-associated changes in mammary gland development.

The suggestion that IGF1 may be the downstream effector of GH signalling in the mammary epithelium originates from observations that *Igf1* messenger RNA (mRNA) is increased in response to GH administration in the mammary epithelium, and that IGF1 can initiate mammary gland differentiation and development in male rats (Ruan et al., 1992). Furthermore, mammary gland development is severely

disrupted in *Igf1*-/- rats, even in the presence of GH (Ruan and Kleinberg, 1999). Administration of IGF1 and oestrogen, but not GH and oestrogen, can rescue the developmental defects in *Igf1*-/- mammary glands, as shown by increased numbers of TEBs, ducts and percent of fat pad occupied (Ruan and Kleinberg, 1999). IGF1 signalling through its cognate receptor IGF1R appears to mediate proliferation in the TEBs, but not apoptosis (Bonnette and Hadsell, 2001). Together, these studies suggest, although they do not conclusively demonstrate, that GH signalling in the mammary stroma acts through IGF1 on the mammary epithelium to drive TEB development and ductal morphogenesis during puberty (Figure 1.3).

Oestrogen signalling

Oestrogen signalling is necessary for ductal elongation and branching in the pubertal mammary gland, and for alveolar differentiation in the pregnant/lactating mammary gland. Oestrogen has two cognate receptors: alpha (ESR1) and beta (ESR2). *Esr1* is expressed in both the mammary epithelium and surrounding stroma (Daniel and Silberstein, 1987). In the absence of ESR1, mammary gland development halts at puberty with no further ductal or alveolar development (Bocchinfuso et al., 2000; Mallepell et al., 2006; Feng et al., 2007). To investigate whether the developmental defects observed in the absence of ESR1 are intrinsic to the mammary epithelium, Mallepell transplanted mammary epithelia from *Esr1*-/- mice and from wild type littermates into contralateral cleared fat pads of a single wild type female recipient, and assessed the ability of the glands to develop (Mallepell et al., 2006). The transplanted wild type glands developed normally and filled the recipient mammary fat pad within 10 weeks after transplantation. During pregnancy, transplanted wild type mammary glands formed additional side-branches and the alveoli differentiated into milk producing cells. On the other hand, the transplanted *Esr1*-/- mammary epithelia failed to develop, and were unable to form differentiated alveolar cells during pregnancy. Mallepell also investigated the role of ESR1 signalling in the mammary stroma by co-transplanting the fat pad from an *Esr1*-/- mouse and the mammary epithelium from a wild type littermate onto the abdominal muscle of a wild type recipient, and assessing the gland's ability to grow.

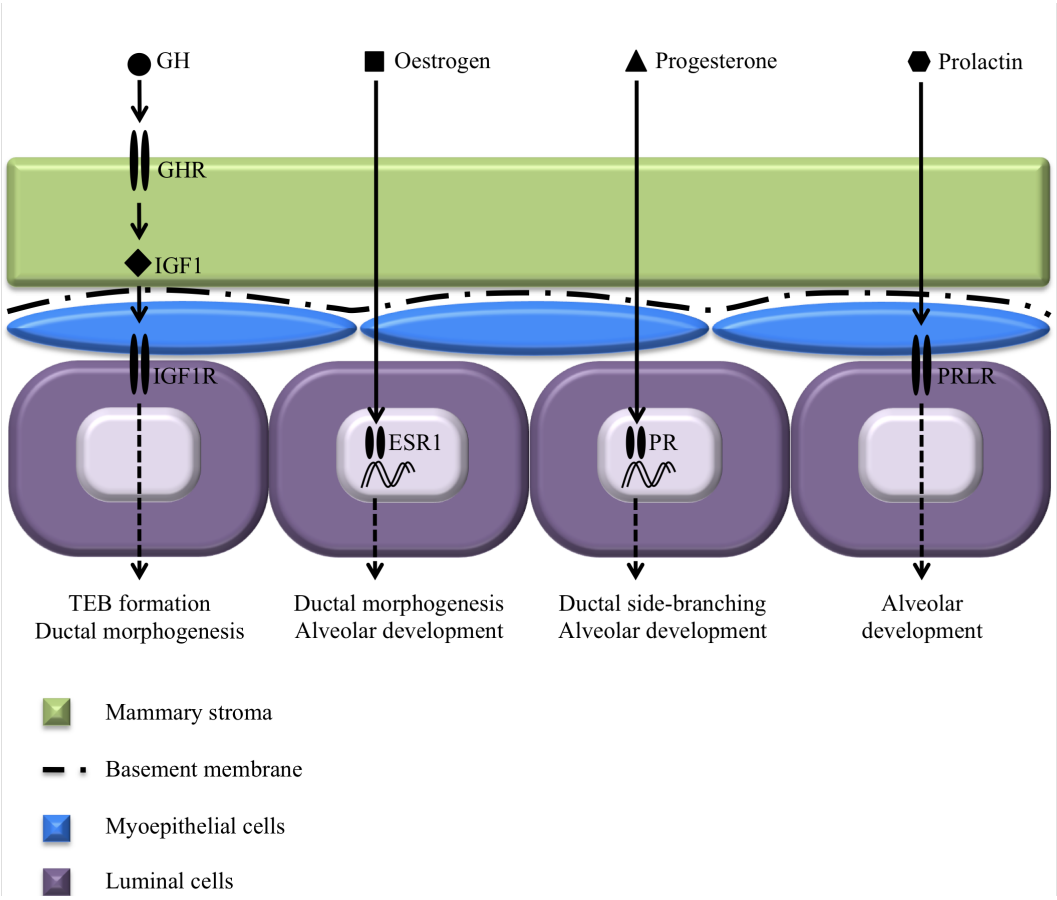


Figure 1.3 Cellular targets of key hormones during postnatal mammary gland development. Activation of growth hormone receptor (GHR) by its ligand (GH) in mammary stromal cells induces the expression of *Igf1*. IGF1, in turn, activates its cognate receptor, IGF1R, in luminal cells, and promotes TEB formation and ductal morphogenesis. Oestrogen activates ESR1 in luminal cells, and regulates ductal elongation and branching during puberty, and alveolar differentiation during pregnancy/lactation. Progesterone receptor (PR) is activated upon binding of its ligand, progesterone, in luminal cells, and promotes ductal side-branching and alveolar development during pregnancy. Prolactin activates its receptor (PRLR) in luminal cells, and controls the proliferation, terminal differentiation and survival of alveolar cells during pregnancy and lactation. **Abbreviations:** GH, growth hormone; GHR, growth hormone receptor; IGF1, insulin-like growth factor 1; IGF1R, insulin-like growth factor 1 receptor; TEB, terminal end bud; ESR1, oestrogen receptor alpha; PR, progesterone receptor; Prlr; prolactin receptor.

The transplanted wild type mammary epithelium grew normally in *Esr1*-/- stroma, and showed alveolar differentiation during pregnancy (Mallepell et al., 2006). Cre-mediated excision of ESR1 from the mammary epithelium further demonstrated that defects in ductal morphogenesis (MMTV-Cre) and alveolar differentiation (WAP-Cre) were intrinsic to the epithelium itself (Feng et al., 2007). As well, through the co-transplantation of chimeric mammary epithelium, which contained both *Esr1*-/- and wild type cells, Mallepell demonstrated that *Esr1*-/- epithelial cells are able to contribute to all parts of the mammary tree, including TEBs, ducts and side-branches, suggesting that paracrine signals from the subset of oestrogen responsive WT epithelial cells are sufficient for normal mammary gland development (Mallepell et al., 2006). Taken together, these data demonstrate the importance of ESR1-mediated regulation of mammary gland development. ESR2, on the other hand, does not appear to be important in mammary gland development. The mammary glands of *Esr2*-/- females develop normally, albeit with a slight delay in the onset of pubertal ductal branching, which has been attributed to decreased progesterone levels in these mice (Antal et al., 2008).

Progesterone signalling

Pregnancy-induced ductal side-branching and alveolar development are regulated in part by progesterone signalling. Progesterone acts through two cognate receptors, PRA and PRB, which are expressed from a single gene (*Pgr*) through two alternate transcription start sites (Kastner et al., 1990; Brisken et al., 1998). Both receptors are expressed in the mammary gland, with PRA more prominently expressed in the virgin, while PRB more prominently expressed in the pregnant mammary gland (Conneely et al., 2003; Aupperlee, 2005; Brisken and O'Malley, 2010). In the absence of PR, the mammary gland develops normally in the virgin mouse. However, while wild type mammary glands form additional branches and develop distinct alveolar structures in response to injected oestrogen and progesterone (used to mimic the increase in these hormones during pregnancy), the mammary glands of PR null mice are unable to do so (Lydon et al., 1995).

To examine whether the phenotype observed in *Pgr*-/- mice is intrinsic to the mammary epithelium or due, at least in part, to other reproductive defects observed in these mice, such as failure to ovulate, Brisken transplanted PR null mammary epithelia into wild type mammary fat pads and conversely, PR null fat pads with normal mammary epithelium into wild type female recipients (Brisken et al., 1998). Through these experiments they demonstrated that PR in the mammary epithelium, but not in the stroma, is required for pregnancy-induced side-branching and the development of differentiated alveolar cells. As well, through the co-transplantation of chimeric mammary epithelium, which contained both *Pgr*-/- and wild type cells, they demonstrated that *Pgr*-/- epithelial cells are able to proliferate and differentiate, suggesting that paracrine signals from a subset of epithelial cells are sufficient for normal mammary gland development (Brisken et al., 1998). Mulac-Jericevic later demonstrated that receptor activator of nuclear factor (NF) κβ ligand (RANKL), an important regulator of osteoclast development, is one such paracrine mediator of progesterone signalling (Mulac-Jericevic et al., 2003; Fernandez-Valdivia et al., 2009). WNT4A was demonstrated to be another paracrine mediator, albeit its involvement is in progesterone-driven branching morphogenesis rather than alveologenesis (Brisken et al., 2000).

To determine which receptor isoform is responsible for progesterone responsiveness in the mammary epithelium, mouse models were generated to selectively ablate either PRA or PRB expression (Mulac-Jericevic et al., 2000; 2003). The *Pgrb*-/- mouse displayed defects in pregnancy-associated ductal side-branching and alveolar development, whereas the mammary glands of *Pgra*-/- mouse developed normally. These data demonstrate that progesterone signalling through PRB is essential for branching and alveolar development during pregnancy.

Prolactin signalling

Prolactin (PRL) signalling is essential for the proliferation, terminal differentiation and survival of alveolar cells in the mammary gland during pregnancy and lactation. Prolactin mediates its signalling through its cognate receptor, PRLR (Boutin et al.,

1988). PRLR has three isoforms in mouse, which are splice-variants transcribed from a single gene (*Prlr*) (Davis and Linzer, 1989; Bole-Feysot et al., 1998). Several studies investigated the role of prolactin signalling in mammary gland development using the PRLR knockout mouse (Ormandy et al., 1997a; 1997b; Brisken et al., 1999; Gallego et al., 2001). The absence of PRLR does not affect development of the mammary gland before puberty. After puberty, however, absence of PRLR leads to perturbed ductal development, with smaller ducts and fewer branch points, and absent alveolar development as evidenced by the persistence of TEBs into adulthood (Ormandy et al., 1997a; 1997b). Ormandy's studies were unable to assess lactation in *Prlr*-/- mice because these females are sterile, while *Prlr*+/- mice, which are not sterile, have severe defects in lactation (Ormandy et al., 1997b). Later studies assessed whether the developmental defects in *Prlr*-/- mice are intrinsic to the mammary epithelium or the result of systemic endocrine defects by transplanting *Prlr*-/- mammary epithelia into the cleared mammary fat pads of wild type recipients (Brisken et al., 1999). In the presence of wild type stroma, *Prlr*-/- mammary epithelia develop normally at puberty, suggesting that the aberrant side-branching observed in *Prlr*-/- mice *in vivo* is an indirect effect of prolactin signalling (Brisken et al., 1999). During pregnancy, however, *Prlr*-/- mammary epithelia fail to develop differentiated alveolar structures, demonstrating that the defects observed in lactation are the result of direct prolactin signalling to the mammary epithelium (Brisken et al., 1999; Gallego et al., 2001).

One key effector of prolactin signalling during pregnancy and lactation is the E74-like factor 5 (ELF5) (Oakes et al., 2008). In *Prlr*-/- mice, *Elf5* expression is down-regulated, whereas in *Elf5*-/- mice, *Prlr* expression is normal, suggesting that PRLR is upstream of ELF5 signalling (Zhou et al., 2005). Signal transducer and activator 5 (*Stat5*) expression is also reduced in *Elf5*-/- mammary glands, and ELF5 directly binds and regulates the *Stat5* promoter (Choi et al., 2009). In the *Elf5*-/- mammary gland, alveolar cells do not differentiate (the phenotype is very similar to the *Prlr*-/- mouse), and conversely, upon over-expression of *Elf5*, precocious alveolar differentiation and milk secretion is observed (Oakes et al., 2008). As well, re-expression of *Elf5* in *Prlr*-/- mammary epithelia restored normal alveolar

development and milk production (Harris et al., 2006). Together, these data demonstrate that prolactin-mediated expression of *Elf5* specifies the development and differentiation of the luminal secretory cells in the mammary gland.

Oestrous cycle and the mammary gland

From the onset of puberty, the cyclical expression of oestrogen and progesterone control the development and maintenance of the virgin mammary gland. (Figure 1.4) The morphological changes that accompany each oestrous cycle (alveologenesis and secondary branching) are driven by the proliferation and apoptosis of a subset of mammary epithelial cells, particularly in the alveolar structures (Ankrapp et al., 1998; Fata et al., 2001a). The highest level of proliferation occurs during late proestrous and oestrous (Andres and Strange, 1999), when ESR1 expression is highest in the mammary epithelium. Oestrogen promotes the elongation of the ducts, and induces the expression of PR in the mammary epithelium (Haslam and Shyamala, 1979), which in turn promotes alveolar development (Brisken et al., 1998). Epidermal growth factor (EGF) induces the expression of PR through activation of oestrogen signalling, thus providing a possible mechanism for initiating the proliferation observed during oestrous, and fundamentally, for the development of hormone responsiveness in the mammary epithelium at the onset of puberty (Ankrapp et al., 1998).

In the absence of pregnancy, the levels of ovarian hormones drop and diestrous (menses) proceeds. The drop in oestrogen and progesterone levels is closely correlated with a peak in apoptosis in the mammary gland, particularly in the alveoli (Andres and Strange, 1999). This process requires degradation and restructuring of the epithelial and stromal networks, which is carried out primarily by matrix metalloproteinases (MMPs) and their inhibitors (TIMPs), whose expression is closely correlated with the cyclical turnover of mammary epithelial cells during the oestrous cycle (Ankrapp et al., 1998; Fata et al., 2001a). MMP2 and MMP3 have been shown to be required for TEB invasion of the fat pad and branching morphogenesis (Wiseman et al., 2003). *Mmp2*-/- mammary glands are defective in

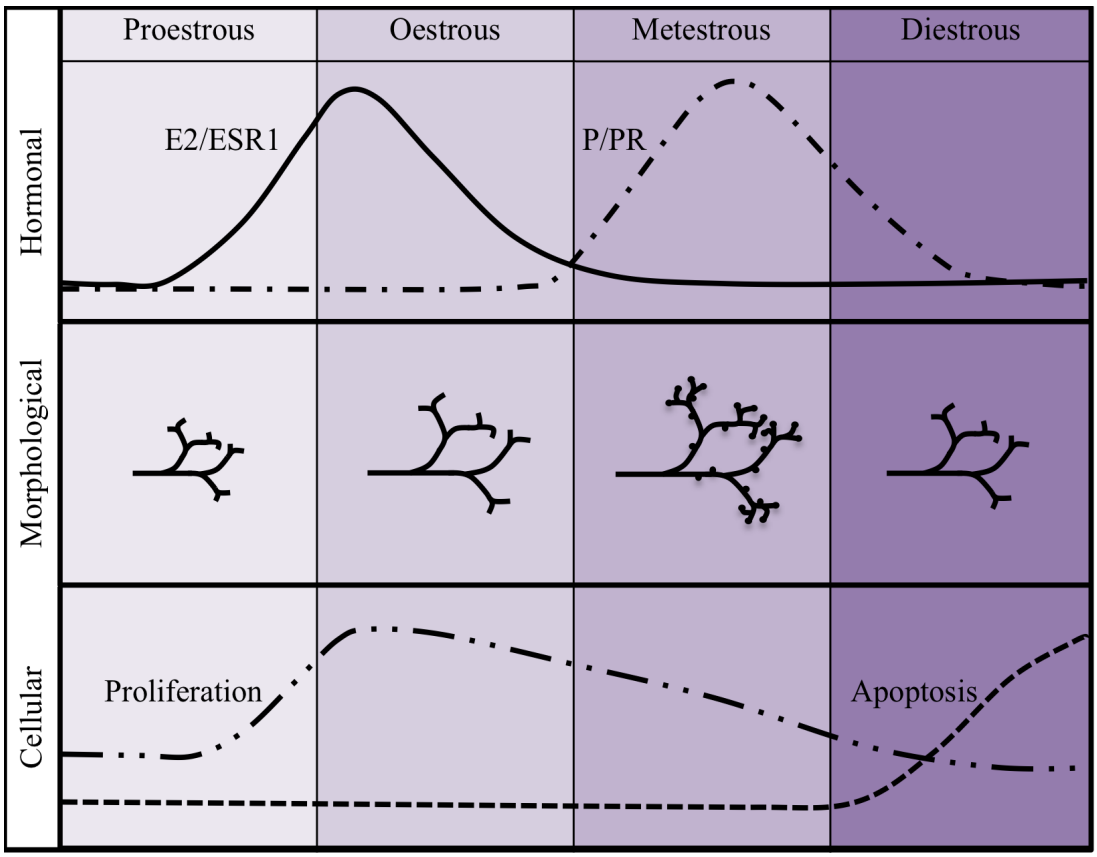


Figure 1.4 Oestrous cycle and the mammary gland. From the onset of puberty, the cyclical expression of oestrogen/oestrogen receptor and progesterone/progesterone receptor (top row) control the development and maintenance of the virgin mammary gland. The morphological changes of alveologenesis and secondary branching that accompany each oestrous cycle (middle row) are driven by the proliferation and apoptosis of a subset of mammary epithelial cells (bottom panel), particularly in the alveolar structures. The highest level of proliferation occurs during late proestrous and oestrous, when ESR1 expression is highest in the mammary epithelium. Oestrogen promotes the elongation of the ducts, and induces the expression of PR in the mammary epithelium, which in turn promotes alveolar development. **Abbreviations:** E2, oestrogen; ESR1, oestrogen receptor alpha; P, progesterone; PR, progesterone receptor.

TEB invasion and have excess secondary branches, whereas MMP3 null mammary glands are defective in secondary branching, but have normal TEB invasion. Inhibition of MMPs by TIMP1 leads to defective TEB invasion, but has little effect on secondary branching (Wiseman et al., 2003), whereas *Timp3*-/- mammary glands display accelerated luminal collapse and loss of alveolar structures (Fata et al., 2001b). These data demonstrate that tightly controlled spatial and temporal expression of ovarian hormones, as well as specific MMPs and TIMPs, are required for the cyclical remodelling of the virgin mammary gland to maintain tissue homeostasis.

1.1.3 Pregnancy, lactation and involution

Overview

With each pregnancy, the mammary gland undergoes a cycle of proliferation, differentiation and apoptosis. The ducts form additional side-branches in response to progesterone and prolactin. These side-branches contain alveolar buds, which functionally differentiate into secretory cells capable of milk production during pregnancy (Robinson et al., 1995). At parturition, prolactin induces the alveoli to produce milk, and suckling stimulates the release of oxytocin, causing the myoepithelial cells surrounding the alveoli to contract, and expel the milk into the primary ducts connected to the nipple. Significant structural remodeling, known as involution, is initiated upon the termination of suckling. Over a 2-week period, the alveoli collapse as a result of apoptotic cell death, and the mammary gland returns to a state that resembles a mature, virgin gland.

Epithelial proliferation during early pregnancy

The first phase of epithelial proliferation and differentiation is initiated early in pregnancy by the synergistic actions of progesterone and prolactin (Brisken, 2002). As described in the previous section, not all luminal epithelial cells are responsive to hormonal cues, therefore mammary epithelial proliferation during early pregnancy is

mediated in part by paracrine signals (Brisken et al., 1998; 2000; Mulac-Jericevic et al., 2003). One such mediator of progesterone (and PRL) signalling is RANKL, which activates the NF-κβ pathway in neighbouring epithelial cells, and in turn activates cyclin D1 (*Ccnd1*) expression (Cao et al., 2001; Srivastava et al., 2003). CCND1 is an important regulator of the cell cycle, and is critical for pregnancy-associated proliferation in normal mammary gland development. In its absence, proliferation is blocked in a cell-autonomous way, particularly in alveoli (Sicinski et al., 1995; Fantl et al., 1999), whereas over-expression of *Ccnd1* is implicated in the proliferation in mammary tumors (Jesselsohn et al., 2010). Another paracrine factor that is important for cell proliferation during early pregnancy is WNT4 (Brisken et al., 2000). Transplantation of *Wnt4*-/- mammary epithelium into wild type stroma demonstrated that in the absence of WNT4, ductal side-branching during pregnancy is greatly reduced. However, by parturition, the number of differentiated alveolar structures is indistinguishable between *Wnt4*-/- and wild type littermates, demonstrating that WNT4 is required by the ductal epithelium during a short time window in early pregnancy.

Polarization of alveolar cells

During mid to late pregnancy, alveolar cells differentiate into polarized pre-secretory cells, and start to highly express genes associated with milk and lipid synthesis (Richert et al., 2000; Anderson et al., 2007). Activation of STAT5 through PRL/JAK2 and ERBB4 is primarily responsible for this transition (Han et al., 1997; Liu et al., 1997; Hennighausen and Robinson, 2005). STAT5 was originally identified as a prolactin-responsive transcription factor responsible for the activation of milk protein genes, such as whey acidic protein (*Wap*) and beta casein (*Csn2*) (Wakao et al., 1994). Two highly homologous isoforms exist, STAT5A and STAT5B (Teglund et al., 1998). *Stat5a* is expressed more highly in the mammary gland, and it is this isoform that is biologically relevant in mammary gland development, as its absence blocks alveolar differentiation into milk-producing cells (Liu et al., 1997). Similar defects in mammary gland development exist upon deletion of *Prlr*, *Jak2*, *Erbb4*, or *Stat5*, underscoring the importance of these

signalling cascades in alveolar cell proliferation and differentiation during pregnancy (Ormandy et al., 1997b; Shillingford et al., 2002b; Long et al., 2003).

Cell adhesion molecules, such as connexins and claudins, are also important in the polarization of alveolar cells, as they regulate and stabilize interactions between the epithelial cells (Streuli et al., 1995; Monaghan and Moss, 1996; Ormandy et al., 2003). Connexins are a component of gap junctions, whose function is to regulate the exchange of small ions and metabolites between cells (Ormandy et al., 2003). Connexin 26 (*Cx26*), a direct STAT5 target gene, is upregulated during pregnancy and lactation in alveolar cells, and loss of *Cx26* (using Cre-loxP based recombination) during early pregnancy leads to increased alveolar cell death (Tu et al., 1998; Ormandy et al., 2003; Bry et al., 2004). Also important for alveolar differentiation are claudins, a major component of tight junctions, which function as barriers to solute transport and the diffusion of molecules, thus helping to maintain cell polarity (Ormandy et al., 2003). In mid to late pregnancy, tight junctions between the luminal epithelial cells close in response to decreased levels of progesterone and increased levels of prolactin (Nguyen et al., 2001). Prolactin signalling not only facilitates closure of tight junctions during pregnancy, but also directs their formation by regulating transcription of their components, such as Claudin-3 and -7 (Ormandy et al., 2003). Closure of the tight junctions prepares the mammary gland for the transport of milk towards the lumen, rather than towards the basement membrane, during lactation.

Interactions with the basement membrane are also required for complete alveolar differentiation (Streuli and Bissell, 1990; Fata et al., 2004). Within each alveolar structure, the luminal cells are surrounded by a discontinuous layer of myoepithelial cells, which allows some of the luminal cells to make direct contact with the basement membrane (Streuli and Bissell, 1990). The basement membrane is created principally from myoepithelial secretion of fibronectin, laminin, collagen, and heparan sulfate (Barcellos-Hoff et al., 1989; Streuli and Bissell, 1990; Dickson and Warburton, 1992; Adams and Watt, 1993). Epithelial interactions with the basement membrane are dependent on β1 integrin (cluster of differentiation [CD] 29), which

dimerizes with alpha integrins to form cell surface receptors for laminins and collagens (Zutter et al., 1998; Fata et al., 2004). In the absence of β1 integrin, alveolar cells become disorganized due to alterations in their associations with the basement membrane, including invasion into the lumen by the epithelium, and proliferation of luminal cells is significantly reduced (Li et al., 2005). Conditional deletion of β1 integrin during early pregnancy demonstrated that it is important in the formation of alveoli, whereas deletion in late pregnancy showed that it is also required for functional differentiation of alveolar cells (Naylor et al., 2005).

Lactation

Secretory differentiation of polarized alveolar cells occurs at parturition as a result of a rapid drop in circulating progesterone and concomitant increase in prolactin (Nguyen et al., 2001; Anderson et al., 2007). High prolactin leads to activation of STAT5-mediated transcription of milk protein genes. However, active lactation is not initiated until progesterone levels are sufficiently low (Kuhn, 1969; Neville et al., 2002). Secretory activation is also associated with a dramatic increase in the cytoplasmic machinery required for lipid and protein synthesis, as well as the disappearance of large lipid droplets that accumulated during pregnancy (Mather and Keenan, 1998; Anderson et al., 2007). Secretion of lipid droplets, which contain milk proteins, is essential for lactation, and defects in its components, for example xanthine dehydrogenase (*Xdh*), lead to accumulation of milk in the luminal cells and the inability to lactate (Vorbach et al., 2002).

The myoepithelium that surrounds the alveolar cells is also actively involved in lactation (Haaksma et al., 2011). By contracting in response to suckling-induced oxytocin stimulation, the myoepithelium reduces the diameter of the alveolar lumen, thereby ejecting the milk into the ductules that connect to the nipple (Gimpl and Fahrenholz, 2001). Although oxytocin is also associated with uterine contractions during birth, females that lack oxytocin have normal gestation and parturition (Nishimori et al., 1996). As well, although oxytocin null females display normal maternal behaviours, all pups die shortly after birth due to the mother's inability to

nurse, demonstrating the essential role of oxytocin in milk ejection during lactation (Nishimori et al., 1996).

Lactation occurs in two phases: the colostral phase and the mature secretion phase (Anderson et al., 2007). During the colostral phase, which is short-lived in the mouse, the milk contains high levels of immunoglobulins and other immune defense proteins, which provide passive immunity for the newborn against pathogens that the mother has been exposed to, and facilitates the establishment of the gut flora (Kelleher and Lönnerdal, 2001; Anderson et al., 2007). During the mature secretion phase, which lasts approximately 3 weeks, milk contains high levels of proteins (12 %), lipids (30 %), and lactose (5 %), which provide the nutrients required to support rapid growth (Anderson et al., 2007). Lactose is a disaccharide unique to milk, and its synthesis is considered a marker for secretory activation of the alveoli (Anderson et al., 2007). In the absence of alpha lactalbumin, an essential cofactor for lactose synthesis, lactation is disrupted and mice cannot sustain their offspring (Stacey et al., 1995).

Microarray studies investigating the temporal changes in gene expression from pregnancy to involution led to the development of a model whereby the activation of metabolic genes required for milk synthesis and secretion is achieved, in part, through prolactin-mediated activation of AKT1 (Rudolph et al., 2003). AKT1 is a serine-threonine kinase that is highly up-regulated by the PRLR/JAK2/STAT5 pathway in late pregnancy (Creamer et al., 2010), and has been shown to be an important regulator of the metabolic pathways involved in milk synthesis (Boxer et al., 2006; Maroulakou et al., 2008). The absence of AKT1 leads to defects in milk production, which result in reduced pup weight and increased pup mortality (Boxer et al., 2006). In addition to its metabolic functions, AKT1 is also a potent anti-apoptotic factor, and its absence is marked by accelerated post-lactational involution (Maroulakou et al., 2008). The exact mechanism by which AKT1 regulates lipid biosynthesis remains unclear. This pathway has been linked to the regulation of the sterol regulatory element binding protein 1 (SREBP1), which is an important regulator of genes involved in fatty acid and cholesterol biosynthesis (Anderson et

al., 2007). However, the direct targeting of SREBP1 by AKT1 signalling has not been demonstrated convincingly, as the expression of *Srebp1* is unchanged in *Akt1*-/- mice (Boxer et al., 2006).

Involution

The cessation of suckling initiates rapid remodelling in the mammary gland, known as involution. Most of the studies investigating this process used forced weaning of the pups (removal of the pups after a specified number of days of suckling), rather than natural weaning, which would occur more slowly (Richert et al., 2000). It is therefore assumed that forced weaning recapitulates natural weaning, but in a controlled manner that allows for specific time points to be assessed (Richert et al., 2000). During the first two days after sucking stops (post-suckling), milk accumulates in the alveoli and the levels of lactogenic hormones begin to fall (Li et al., 1997). The expression of milk protein genes drops in response to reduced levels of prolactin; however, lactation can be reinitiated with suckling during this time period (Li et al., 1997; Marti et al., 1999). To identify the mechanisms that regulate this initial stage of involution, Li used three models (sealing of teats, mammary gland transplants, and inactivation of oxytocin) to disrupt milk secretion while maintaining normal systemic hormonal regulation (Li et al., 1997). These studies demonstrated that local factors, rather then systemic hormones, are required for the induction of programmed cell death that result primarily from changes in the phosphorylation of STAT3 (Li et al., 1997). While STAT5 was required for alveolar development and differentiation, STAT3 is essential for the initiation of apoptosis during involution (Chapman et al., 1999; Humphreys et al., 2002). Leukemia inhibitory factor (LIF) and transforming growth factor beta 3 (TGFβ3) activate *Stat3* (Nguyen and Pollard, 2000; Kritikou et al., 2003; Schere-Levy et al., 2003), which in turn activates the transcription of key pro-apoptotic genes, such as *Igfbp5* and *Cebpd* (Tonner et al., 2002; Thangaraju et al., 2005). Apoptosis can be suppressed by AKT/protein kinase B (PKB), a cell survival complex that is negatively regulated by STAT3-mediated expression of its inhibitor, phosphatidylinositol 3 kinase (PI3K) (Ormandy et al., 1997b; Schwertfeger et al., 2001; Shillingford et al., 2002b; Long et

al., 2003; Abell et al., 2005). Another pathway that is important for regulating apoptosis during involution is the NF-κβ/Iκβ kinase (IKK)- death receptor (DR) pathway. Upon conditional deletion of *Ikk* from the mammary gland, a delay in apoptosis and remodelling is observed, as well as reduced expression of NF-κβ targets, such as tumour necrosis factor (*Tnf*) and its cognate receptor (*Tnfr1*) (Sakakura, 1987; Ormandy et al., 2003; Baxter et al., 2006).

Removal of the large number of apoptotic epithelial cells from the mammary lumen is achieved through efferocytosis (phagocytosis of apoptotic cells) by both epithelial cells and macrophages (Tu et al., 1998; Ormandy et al., 2003; Bry et al., 2004; Monks et al., 2005; 2008; O'Brien et al., 2012). Monks observed that viable epithelial cells engulfed nearby apoptotic cells using many of the same receptors as macrophages, including the cell surface receptors CD36, CD91 and the vitronectin receptor (Monks et al., 2005), as well as the phosphatidylserine receptor (PSR), the activation of which causes rapid cytoskeletal reorganization and macropinosis (Hoffmann et al., 2001). Macrophages also participate in the rapid clearance of apoptotic epithelial cells. Conditional deletion of macrophage colony stimulating factor 1 receptor (*Mcsf1r*) leads to delayed involution, as well as lack of alveolar regression or adipocyte repopulation (O'Brien et al., 2012). Injection of M2-polarized macrophages into *Mcsf1r*-/- mammary epithelium was sufficient to rescue these defects, demonstrating their importance in remodelling the postpartum mammary gland (O'Brien et al., 2012).

1.2 Mouse mammary gland stem and progenitor cells

1.2.1 Stem cells in the mammary gland

Foundations for the mammary stem cell hypothesis

The ability of the mammary gland to undergo repeated cycles of proliferation, differentiation and involution suggests that mammary epithelial stem cells exist. In the 1950s and ‘60s, DeOme and Daniel’s seminal work showed that the serial

transplantation of epithelial fragments isolated from different areas of the mammary gland was able to reconstitute a complete and functional mammary tree in recipient mice (Deome et al., 1959; Daniel et al., 1968). They developed a technique of de-epithelializing the recipient mammary stroma prior to transplantation, yielding a cleared mammary fat pad, which has become a cornerstone assay in the field of mammary gland biology. Subsequent experiments using mammary gland tissue from young and old female mice further demonstrated that the capacity to regenerate a functional mammary gland is maintained throughout life (Young et al., 1971). The identity of mammary stem cell(s) remained elusive because transplantation of any portion of the mammary tree from any aged mouse during any developmental stage gives rise to functional mammary epithelial outgrowths (Smith and Medina, 1988). To determine whether a single cell could generate a functional mammary gland, Kordon and Smith transplanted MMTV-infected cells into cleared fat pads, and assessed the number of integration sites in each outgrowth using Southern blotting (Kordon and Smith, 1998). MMTV is a murine milk-transmitted retrovirus that inserts DNA copies of its genome randomly into the host genome, but expresses itself only in the mammary epithelium (Bittner, 1936; Matsuzawa et al., 1995). Kordon and Smith hypothesized that if an outgrowth originated from a single cell, then only one pattern of integration would be observed through Southern blotting from different regions of the MMTV-infected outgrowth. However, if an outgrowth originated from several cells, then different patterns of integration would be observed. Their observation that each outgrowth contained only one pattern of viral genome integration supported the hypothesis that a single cell could give rise to a functional mammary gland. Through serial transplantation experiments, they further demonstrated that these cells had the potential to self-renew, as the secondary outgrowths were of the same clonal origin as the first (Kordon and Smith, 1998). Although these experiments strongly suggested that mammary stem cells exist, they did not prove that a single cell gave rise to a complete and functional organ, because one could not rule out that a small number of lineage-restricted progenitor cells together gave rise to the different cell types observed in the outgrowths.

The mammary gland side population

Based on recent discoveries concerning other tissue stem cells, such as those in the hematopoietic system (Goodell et al., 1996), Welm and Alvi identified and isolated the first putative mammary stem cell population, termed mammary gland side population (MG-SP) (Welm et al., 2002; Alvi et al., 2003). SP cells are identifiable through flow cytometry due to their ability to expel Hoechst 3342 dye, thereby forming a distinct population of cells off to the side of the main cell population (hence the term 'side population') (Goodell et al., 1996). In the hematopoietic system, SP cells have greater than 1000-fold higher *in vivo* reconstitution activity than non-SP cells, and are multi-lineage, as demonstrated by their contribution to both myeloid and lymphoid lineages (Goodell et al., 1996). However, the use of Hoechst dye to identify stem cells is problematic because Hoechst dye interferes with DNA replication during cell division. Therefore, it is possible that the difference observed in the *in vivo* reconstitution activity between hematopoietic SP and non-SP cells is due to Hoechst toxicity to the cells that are unable to efflux the dye, rather than a biological difference in their repopulating potential. In the mammary gland, SP cells are bromo deoxyuridine (BrdU)-label retaining and form approximately 3 % of the mammary epithelium (Welm et al., 2002). It was first assumed that cells retaining BrdU are quiescent (Welm et al., 2002), but a later study by Smith demonstrated that a high percentage (~83 %) of BrdU retaining cells are actively dividing through asymmetric cell divisions, which leads to their label retaining property (Smith, 2005). Upon *in vitro* culturing, MG-SP cells form epithelial colonies that express classical mammary epithelial cell markers, such as K14, K18 and K19. Upon *in vivo* transplantation at limiting dilutions, MG-SP cells give rise to both ductal and alveolar structures, albeit with varied frequency between different research groups (Welm et al., 2002; Alvi et al., 2003; Liu et al., 2004). MG-SP cells also express cell surface markers that are associated with stem cell activity, such as stem cell antigen 1 [Sca1; (Spangrude et al., 1988; Welm et al., 2002)] and alpha-6 integrin [CD49f; (Li et al., 1998; Stingl et al., 2006)], and are responsive to growth factors, such as WNT3A and EGF (Liu et al., 2004).

However, later studies (reviewed in the next section) revealed that Sca1 is not expressed in a multipotent stem cell in the mammary gland (Shackleton et al., 2006; Stingl et al., 2006), but is highly expressed in at least one progenitor population (Sleeman et al., 2007; Li et al., 2009). Together, these data demonstrate that the SP phenotype is associated with stem cell activity, however, since non-SP cells are also able to reconstitute a cleared fat pad upon transplantation, albeit with variable efficiency, additional markers are needed to specifically label mammary stem cells.

Transplantation studies identify a multipotent mammary stem cell

Another population of mammary stem cells was identified by three independent groups (Shackleton et al., 2006; Sleeman et al., 2006; Stingl et al., 2006). Sleeman first demonstrated that the cell surface antigen CD24 can be used to isolate three cell populations in the mammary gland: $CD24^{negative\ (neg)}$, $CD24^{low\ (lo)}$ and $CD24^{high\ (hi)}$, marking non-epithelial, myoepithelial, and luminal epithelial compartments based on keratin expression, respectively (Sleeman et al., 2006). Transplantation experiments demonstrated that the $CD24^{lo}$ compartment contains the highest repopulating activity, with lower repopulating activity observed in the $CD24^{hi}$ population, suggesting that the mammary stem cell has a basal (myoepithelial-like) position in the cellular hierarchy.

Stingl further refined the purification of the mammary stem cell, which they termed mammary repopulating unit or MRU, with the inclusion of additional cell surface markers (Stingl et al., 2006). By first using negative selection to remove unwanted cell lineages (Lin: CD31-endothelial & CD140a-mesencymal cells; CD45 & Ter119-hematopoietic cells), and then by assessing CD24 and CD49f expression, Stingl isolated three distinct cell populations, and determined that the MRUs were enriched in the population with the cell surface phenotype: $Lin^{neg}CD24^{lo}CD49f^{hi}$. These cells also expressed K5, K14, smooth muscle actin (SMA), smooth muscle myosin, vimentin and laminin, all of which are characteristic markers of the basal/myoepithelial lineage. Transplantation of a mixture of green fluorescent protein (GFP) and cyan fluorescent protein (CFP)-labeled MRUs generated single-colored

outgrowths, suggesting that a single stem cell, rather than several lineage restricted progenitor cells, gave rise to each gland. This was formally demonstrated by performing single cell transplantation of the MRU, which allowed Stingl to determine its frequency to be 1:1400 total epithelial cells, or 1:200 in the $Lin^{neg}CD24^{lo}CD49f^{hi}$ population. These outgrowths contained both myoepithelial and luminal cells, demonstrating that the MRU has multi-lineage potential. Over 90 % of MRUs were in the G1 or S/G2/M phase of the cell cycle, suggesting that these cells are actively dividing. As well, less than 10 % of the MRUs were found in the SP compartment, further confirming that non-SP cells contain a significant proportion of mammary gland stem cells.

Concurrent to this work, Shackleton used another staining strategy to isolate the mammary stem cell (Shackleton et al., 2006). He first selected against the lineage markers CD31, CD45 and Ter119, and then assessed the expression of CD29 (β1 integrin) and CD24 (Shackleton et al., 2006). CD29 was chosen based on its association with stem cell activity in the skin (Jones et al., 1995). This gating strategy allowed for the isolation of four distinct cell populations, each of which was assessed for repopulating activity through transplantation experiments. Repopulating activity was highly enriched in the $Lin^{neg}CD24^{pos}CD29^{hi}$ population, with 1:64 cells giving rise to a new gland. The clonality of the outgrowths was demonstrated through mixing experiments, in which wild type and LacZ-marked cells were co-transplanted. Almost all of the outgrowths (95/97) were either wild type or $LacZ^{pos}$, suggesting that each outgrowth is the result of one transplanted stem cell. Single-cell transplantation experiments confirmed this, and revealed that the stem cell frequency within the $Lin^{neg}CD24^{pos}CD29^{hi}$ population is between 6-12 %. In agreement with Stingl, Shackleton observed high expression of CD49f and K14 in the MRU population with multi-lineage contribution to the outgrowths. Self-renewal of these stem cells was demonstrated by serial transplantations, in which third generation transplants achieved functional outgrowths. Also, the transplantation of a single outgrowth achieved repopulation in several recipients, indicating that the original stem cell that gave rise to the first outgrowth has divided symmetrically to generate more stem cells.

Lineage tracing identifies multipotent and uni-lineage stem cells

Van Keymeulen posited that transplantation studies, although important for defining the differentiation potential of stem cells, may not accurately recapitulate the normal behaviour of stem cells *in vivo*, because in a transplantation setting, stem cells may be forced to differentiate into lineages to which they would not normally contribute (Van Keymeulen et al., 2011). For example, hair follicle bulge stem cells contribute to all epidermal lineages following transplantation, but only to hair follicle regeneration under physiological conditions (Blanpain et al., 2004; Blanpain and Fuchs, 2009). To determine whether the mammary gland is maintained by a multipotent stem cell population, as suggested by transplantation studies, or by lineage-restricted stem or progenitor cells, Van Keymeulen performed lineage-tracing experiments using different keratin reporter lines (K14, K5, K18, K8; generated by using the Cre/loxP system) and assessed the contribution of each marked cell population to the different cell lineages in the mouse mammary gland (Van Keymeulen et al., 2011). In E17 embryos, all mammary epithelial cells (MECs) express K14. At birth, however, K14 expression becomes restricted to the myoepithelial lineage, which also expresses K5 and SMA, whereas the luminal cells are $K14^{neg}$, and instead express K8 and K19. To determine whether embryonic $K14^{pos}$ cells are precursors to all mammary gland cell lineages in the postnatal mouse, including the $K14^{neg}$ luminal cells, K14-expressing cells were labeled with yellow fluorescent protein (YFP) at E17, and their lineage-contribution was assessed in the pubertal mammary gland. YFP^{pos} cells contributed to both myoepithelial and luminal lineages in the embryo, and this multi-lineage contribution was maintained at puberty, suggesting that embryonic $K14^{pos}$ cells are multipotent.

To investigate if postnatal $K14^{pos}$ cells contained multipotent stem cells, lineage tracing was performed at puberty and in adulthood. Postnatal YFP^{pos} cells contributed only to the myoepithelial lineage, and interestingly approximately 30 % were maintained over time, including during pubertal proliferation and after several cycles of pregnancy/involution, suggesting that postnatal $K14^{pos}$ cells contain a long-term uni-lineage stem cell population. To rule out that their strategy targeted a

subpopulation of quiescent myoepithelial cells rather than unilineage stem cells, Van Keymeulen repeated these experiments with a K5 reporter mouse line, and obtained similar results, confirming that the myoepithelial lineage is maintained by a unilineage stem cell population that expresses K14 and K5.

Evidence from Smith's laboratory using the WAP-Cre mouse model demonstrated that alveolar stem/progenitor cells are present in the virgin and lactating mammary gland (Wagner et al., 2002). To assess whether or not these luminal stem cells exist under physiological conditions, K8-expressing cells were marked with YFP at birth, puberty and adulthood, and their contribution to the different lineages was assessed (Van Keymeulen et al., 2011). YFP^{pos} cells only contributed to the luminal lineage, regardless of when during development the marking occurred. Approximately 40 % of YFP^{pos} cells appeared to be mature luminal cells, as they were lost over time. Approximately 4 % of YFP^{pos} cells, however, were able to proliferate and produce mature secretory cells upon pregnancy and lactation, and were maintained over time, including after several cycles of pregnancy/involution, suggesting that these cells are long-term uni-lineage stem cells. To further demonstrate that K8 marked luminal stem cells, and not a subpopulation of mature luminal cells, Van Keymeulen used another luminal marker, K18. Although K18-YFP marked only luminal cells, and YFP^{pos} cells did not proliferate during pregnancy, suggesting that they may be terminally differentiated cells, rather than stem or progenitor cells.

The theory that the mammary gland is maintained by two uni-lineage stem cell populations is inconsistent with the data from transplantation experiments, which demonstrated that a single basal stem cell can give rise to both luminal and myoepithelial lineages. To address this, Van Keymeulen transplanted postnatal K14-expressing (YFP^{pos}) cells into cleared recipient fat pads and showed that these cells were also able to generate a fully functioning mammary gland, with YFP^{pos} cells contributing to both luminal and myoepithelial lineages. However, when K14 (YFP^{pos}) cells were co-transplanted with unlabeled luminal cells, very little YFP contribution was seen in the luminal lineage. When K8 (YFP^{pos}) cells were transplanted alone, they were not able to reconstitute a cleared fat pad; however,

when co-transplanted with unlabeled myoepithelial cells, they generated a functional mammary gland, with YFP contribution observed only in the luminal compartment. These data demonstrate that the conditions used for the transplantation of stem cells can affect their differentiation potential. However, it remains unclear how myoepithelial stem cells can give rise to the luminal lineage. Van Keymeulen hypothesized that either luminal stem cells restrict the differentiation potential of myoepithelial stem cells *in vivo*, or that the differentiation of myoepithelial stem cells towards the luminal fate is a very rare event; a notion that she could not address within her experimental conditions.

1.2.2 Progenitor cells in the mammary gland

Luminal progenitor cells

In the hematopoietic system, the progression from a multipotent stem cell to each of the differentiated cell types occurs through a series of progenitor cells with defined lineage potentials. In the mammary gland, some evidence exists that distinct progenitor cells serve as intermediates between mammary gland stem cells and their differentiated progeny (Asselin-Labat et al., 2006; Shackleton et al., 2006; Stingl et al., 2006; Sleeman et al., 2007; Li et al., 2009). A mammary gland progenitor population (mammary colony forming cell or Ma-CFC) was identified by Stingl and Shackleton (Shackleton et al., 2006; Stingl et al., 2006). These cells have the cell surface phenotype $Lin^{neg}CD24^{hi}CD49f^{lo}$ or $Lin^{neg}CD24^{hi}CD29^{lo}$ and are more numerous than MRUs, with an estimated frequency of 1:63 epithelial cells. They also express luminal cell markers, such as K6, K8, K18, K19 and casein genes. In *in vitro* colony forming assays, these cells form uniform, round structures composed of cuboidal epithelium. In contrast, MRUs form dense, irregularly-shaped colonies under the same culturing conditions. When cultured under lactogenic conditions, 85 % of Ma-CFCs form alveolar-like structure that produce milk proteins, whereas MRUs form a heterogeneous mix of colonies, including branched colonies and spheroid bodies (Shackleton et al., 2006). Inclusion of an additional cell surface marker, CD61 (β3 integrin), allowed for the isolation of a cell population with

increased proliferative activity (Asselin-Labat et al., 2006). Approximately 1:3 $Lin^{neg}CD24^{hi}CD29^{lo}$ cells express CD61. $CD61^{pos}$ cells form larger colonies in colony forming assays, and approximately 10-fold more colonies than $CD61^{neg}$ cells when embedded in Matrigel. The colonies are uniform, round structures, and express K18, a luminal cell marker. *In vivo*, the percentage of $CD61^{pos}$ progenitor cells decreases by approximately 30 % from puberty to adulthood, and by a further 15-fold by day 18 of pregnancy. The progressive loss of progenitor cells during pregnancy is compatible with an increase in the differentiation of secretory cells. A block in the differentiation of luminal cells due to the loss of GATA3 leads to the accumulation of $CD61^{pos}$ cells, providing further evidence that these cells are precursors to the differentiated cell types of the lumen.

Sleeman used two different cell surface markers, Sca1 and prominin1 (CD133), in addition to CD24, to isolate two luminal progenitor populations (Sleeman et al., 2007). Sca1 and prominin1 expression were almost completely overlapping. Both $CD133^{pos}$ and $CD133^{neg}$ cell populations express K18, with the $CD133^{pos}$ population expressing higher levels than the $CD133^{neg}$ population. The $CD133^{pos}$ progenitors also express genes involved in hormone sensing, such as *Esr1*, *Pr* and *Prlr,* while the $CD133^{neg}$ progenitors express genes associated with alveolar function, such as *Csn2* and *Wap*. In colony forming assays, $CD133^{neg}$ cells displayed higher proliferative activity, with more than 40 % of cells plated giving rise to colonies, whereas only 15 % of $CD133^{pos}$ cells yielded colonies. Neither population is efficient in reconstituting cleared mammary fat pads following transplantation. Further subdivision of $Sca1^{pos}$ and $Sca1^{neg}$ progenitor cell populations based on their expression of the cell surface marker alpha 2 integrin (CD49b) also showed that the $CD49b^{pos}$ compartment is enriched in progenitor cell activity (Li et al., 2009).

Myoepithelial progenitor cells

Myoepithelial progenitor cells have thus far been defined as the *in vitro* counterparts of MRUs. Stingl defined the myoepithelial progenitor as $Lin^{neg}CD24^{lo}CD49f^{hi}$; Shackleton as $Lin^{neg}CD24^{lo}CD29^{hi}$; and Sleeman as $CD24^{lo}Sca1^{neg}CD133^{neg}$

(Shackleton et al., 2006; Stingl et al., 2006; Sleeman et al., 2007). In *in vitro* colony forming assays, all three populations generated large, irregular-shaped colonies, and expressed K14 and SMA, markers of the myoepithelium. The frequency of colony forming activity varied from 5 % to 25 % between the studies, as well, some differences were observed between the *in vivo* and *in vitro* potential of the cells. The reasons behind the observation that the $CD24^{lo}$ population displayed high *in vivo* repopulating activity and low *in vitro* colony forming activity in Sleeman's studies was not determined.

1.3 CCAAT enhancer binding proteins

1.3.1 The CCAAT enhancer binding protein family

The first CCAAT enhancer binding protein (C/EBP), later termed C/EBPα, was described as a heat stable protein in rat liver nuclei capable of interacting with the CCAAT box of the herpes virus thymidine kinase promoter (Graves et al., 1986). Subsequent studies identified five additional genes (*Cebpb*, *Cebpd*, *Cebpg*, *Cebpe* and *Cebpz*) that were named in the order of their discovery (Chang et al., 1990; Cao et al., 1991; Descombes and Schibler, 1991; Williams et al., 1991; Ron and Habener, 1992). This family of transcription factors plays key regulatory roles in a number of tissues, including the blood, liver, adipose tissue, brain, and the mammary gland (Descombes and Schibler, 1991; Sterneck et al., 1997; Seagroves et al., 1998; Kowenz-Leutz and Leutz, 1999; Porse et al., 2001; Benavides et al., 2005). C/EBPs share a modular domain architecture with a variable N-terminus and a highly conserved C-terminus (Figure 1.5) (Ramji and Foka, 2002). The N-terminus of each C/EBP contains a unique combination of transactivation domains (AD) and repressor domains (RD), leading to its distinct biological role, which can range from a strong transcriptional activator, such as C/EBPα, to a dominant negative repressor, such as C/EBPγ (Cooper et al., 1995; Nerlov, 2007). The C-termini of C/EBPs encode the DNA-binding (DBD) and dimerization (bZIP) domains, and display greater than 90 % conservation between family members (Ramji and Foka, 2002). The dimerization domain contains a heptad repeat of four or five leucine residues that form an alpha

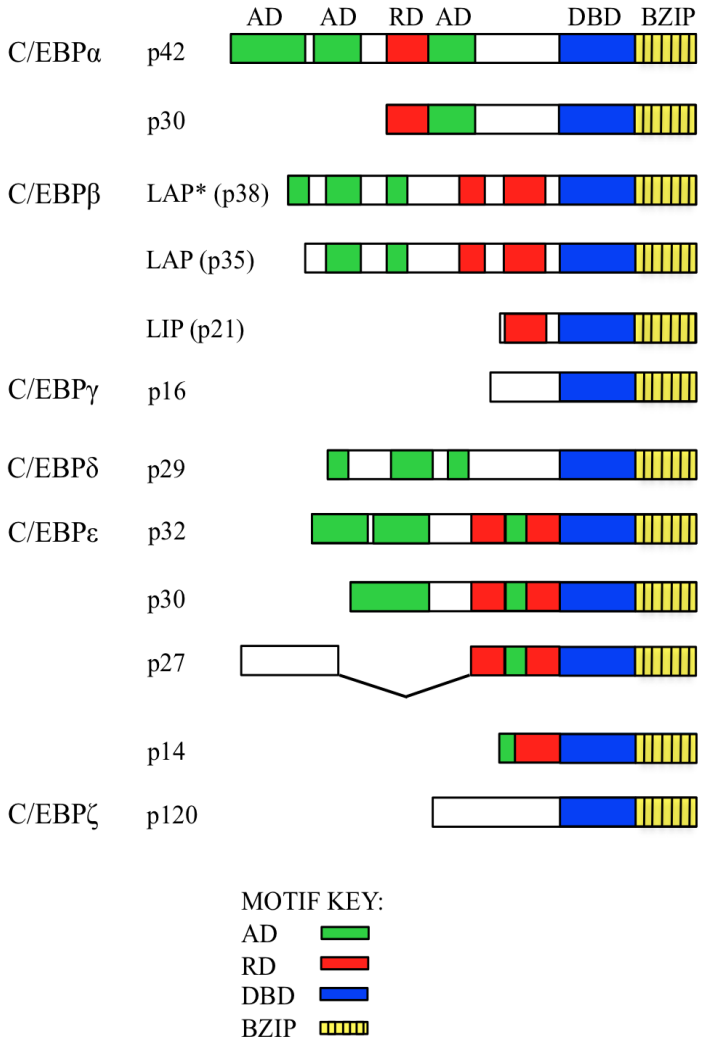


Figure 1.5 The CCAAT enhancer binding protein (C/EBP) family. C/EBPs are transcription factors that have key regulatory roles in diverse cellular processes, such as cell proliferation and differentiation**.** The two isoforms of C/EBPα (p42 and p30) and the three isoforms of C/EBPβ (LAP* [p38], LAP [p35] and LIP [p21]) arise from alternate translation start site from a single RNA transcript for each gene. Whereas the four C/EBPε isoforms arise from differential RNA splicing (p27), alternate promoters (p29, p14 and p33/p30/p27) and alternate translation start sites (p32 & p30). The six C/EBP family members share a modular domain architecture with a variable N-terminus (green and red) and a highly conserved C-terminus (blue and yellow). The N-terminus of each C/EBP contains a unique combination of transactivation domains (AD-green) and repressor domains (RD-red), leading to its distinct biological role, which can range from a strong transcriptional activator, such as C/EBPα, to a dominant negative repressor, such as C/EBPγ. The C-termini of C/EBPs encode the DNA-binding (DBD-blue) and dimerization (BZIP-yellow) domains, and display greater than 90 % conservation between family members. This figure is adapted from Ramji and Foka, 2002.

helix (Landschulz et al., 1989; Vinson et al., 1989). Dimerization, a prerequisite for DNA binding, results from the formation of a coiled-coil structure when two such leucine repeats, each from a C/EBP monomer, come into close proximity (i.e., leucine zipper or bZIP) (Vinson et al., 1989). C/EBPs can form homo- or heterodimers with other C/EBP family members, or other transcription factors, such as C-FOS, cyclic adenosine monophosphate (cAMP) response element binding protein (CREB), and activating transcription factor (ATF) (Lekstrom-Himes and Xanthopoulos, 1998). The amino acid residues in the basic regions of the protein dimer determine the specificity of DNA binding (Osada et al., 1996). The consensus binding sequence for C/EBPs is a dyad symmetrical repeat ($^{A}/_{G}$TTGCG$^{C}/_{T}$AA$^{C}/_{T}$) (Osada et al., 1996). Heterodimers with other bZIP proteins are unable to bind canonical C/EBP sites, and either bind to other regulatory sequences, or sequester and repress C/EBP transcriptional activity (Takiguchi, 1998). C/EBPs are also able to interact with non-bZIP transcription factors and co-activators, such as NF-κβ p50 subunit (LeClair et al., 1992), glucocorticoid receptor [GR, (Nishio et al., 1993)], PR (Christian et al., 2002), and CREB binding protein [CBP, (Mink et al., 1997)]. C/EBPα and C/EBPβ have also been shown to inhibit the cell cycle and promote terminal differentiation through their interaction with the E2F complex (Porse et al., 2001; Sebastian et al., 2005).

1.3.2 Mouse C/EBPβ

Regulation of Cebpb gene expression

The *Cebpb* gene is located on mouse chromosome two, and is comprised of a single exon that is transcribed into a 1.4 kb mRNA. The gene and its promoter have been characterized in human, mouse, rat, chicken and *Xenopus laevis* (Grimm and Rosen, 2003). The expression of *Cebpb*, also known as *Lap* or *Crp2*, is transcriptionally controlled by an auto-feedback loop, as well as by CREB/ATF, C-JUN and SP1 in response to diverse extracellular stimuli (Chang et al., 1995; Niehof et al., 1997; Berrier et al., 1998; Niehof, 2001; Ruffell et al., 2009) (Figure 1.6). For example, in the liver, *Cebpb* expression is upregulated in response to inflammatory stimuli by

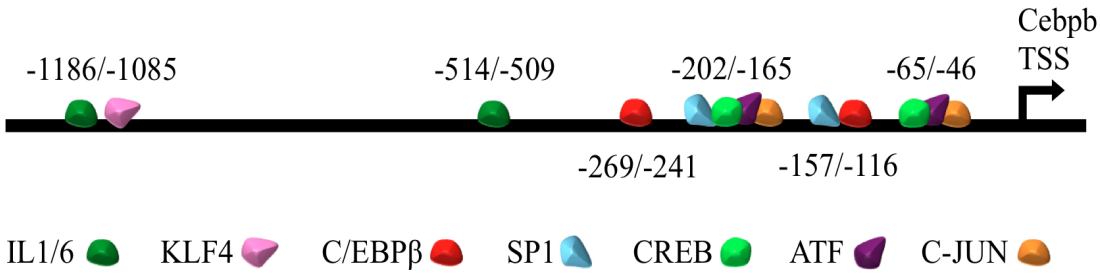


Figure 1.6 Transcriptional regulation of *Cebpb* expression. The expression of *Cebpb* is regulated by a number of factors, including CREB, ATF, C-JUN and SP1, KLF4 and IL1/6. Regulation by some of these factors is tissue specific. For example, IL1 and IL6 regulate *Cebpb* expression in the liver in response to inflammation. *Cebpb* expression in fat cells, on the other hand, is primarily regulated by KLF4. Other factors, such as CREB, ATF, C-JUN, SP1 and C/EBPβ contribute to the regulation of *Cebpb* expression in all tissues.

lipopolysaccharides, interleukins (IL) 1 and 6, gamma interferon and turpentine (Akira et al., 1990; Ray, 1994). In fat cells, the adipogenic response to insulin and glucocorticoids is regulated by Krüppel-like factor 4 (KLF4)-mediated expression of *Cebpb* (Birsoy et al., 2008). In contrast, in the ovaries, *Cebpb* expression is controlled by luteinizing hormone (LH)-mediated EGF/RAS/ERK signalling, the activation of which leads to granulosa cell differentiation (Fan et al., 2009; 2011). The *Cebpb* promoter contains two CREB binding sites that have been shown to mediate *Cebpb* response to cAMP via the protein kinase A (PKA) pathway (Niehof et al., 1997; Ruffell et al., 2009). Hormones and other signals that lead to CREB phosphorylation via the mitogen-activated protein kinase (MAPK) pathway can target these binding sites as well to induce *Cebpb* expression.

C/EBPβ has three protein isoforms

Three different proteins are produced from the single *Cebpb* transcript by using alternate translation initiation codons (Descombes and Schibler, 1991). The full-length isoform, known as liver-enriched transcriptional activator protein (LAP*; 38 kilodaltons [kDa]), is translated from the first AUG codon. LAP* contains three AD and two RD domains. LAP (35 kDa) is translated from the second AUG codon. LAP lacks the first 21 N-terminal amino acids, but retains two of the three ADs and both RD domains. Although almost identical in sequence and structure, LAP* and LAP are functionally distinct (Kowenz-Leutz and Leutz, 1999). For example, only LAP* is able to interact with MYB to induce transcription of the granulocyte-specific gene *Mim1* (Kowenz-Leutz and Leutz, 1999). This specificity may be due to an interaction between the 21 N-terminal amino acids of LAP* and the SWI/SNF complex involved in chromatin remodeling (Kowenz-Leutz and Leutz, 1999; Eaton et al., 2001).

Liver-enriched transcriptional inhibitor protein (LIP; 21kDa) is translated from the third AUG codon, and lacks all three AD and one of the two RD domains, while the bZIP and DBD domains remain intact. Retention of the dimerization and DNA-binding capacity allows LIP to interact with other C/EBPs and bind to target DNA

sequences. However, because of the absence of the AD domains, LIP homo- and heterodimers are unable to initiate transcription, and therefore are considered to be dominant negative repressors (Descombes and Schibler, 1991). The ratio of LAP to LIP also appears to be important, because a moderate increase in this ratio leads to a significant increase in transcriptional activation of target genes (Gomis et al., 2006).

As well, phosphorylation plays a key role in C/EBPβ regulation (Takiguchi, 1998). Several serine and threonine residues are targets of phosphorylation by a number of signalling pathways. C/EBPβ exists in a repressed state because the negative regulatory domains interact with, and thereby inhibit, the transactivation and DNA binding domains (Kowenz-Leutz and Leutz, 1999; Mo et al., 2004). Upon RAS activation, MAPK phosphorylates C/EBPβ (the site of phosphorylation is species-specific) and leads to the release of this intramolecular repression (Kowenz-Leutz et al., 1994). Other kinases known to phosphorylate C/EBPβ include PKA (Trautwein et al., 1994), protein kinase C [PKC, (Trautwein et al., 1993; 1994)], calcium calmodulin-dependent kinase [CaMK, (Wegner et al., 1992)], ribosomal S6 kinase 2 [RSK2, (Xing et al., 1996)] and PI3K (Guo et al., 2001). In addition to relieving the repression of C/EBPβ, phosphorylation is also important for trafficking activated C/EBPβ to the nucleus (Ford et al., 1996).

1.3.3 Expression of C/EBPs in the mammary gland

C/EBPα

Three C/EBP family members are expressed in the mammary gland: C/EBPα, -β, and -δ. *Cebpa* is present at constant levels throughout mammary gland development, except during lactation, when *Cebpa* expression increases (Gigliotti and DeWille, 1998; Robinson et al., 1998; Seagroves et al., 1998). To determine if C/EBPα is important for the development of the mammary epithelium, *Cebpa*-/- mammary glands from newborn mice were transplanted into cleared mammary fat pads and the outgrowths were assessed at different developmental time points (Seagroves et al., 1998). The *Cebpa*-/- outgrowths were comparable to controls, with no overt defects

observed in the virgin, pregnant, lactating or involuting gland. These data suggest that C/EBPα is not required for normal development of the mammary epithelium. However, its role in the mammary stroma, which is primarily composed of adipocytes, a cell type in which C/EBPα is known to serve an important developmental function (Porse et al., 2001), has not been directly investigated. Mice in which C/EBPα is deleted in the liver do not develop white adipose tissue (Linhart et al., 2001). The mammary fat pads, however, appear normal in these mice, suggesting that this adipose tissue may develop independently of C/EBPα.

C/EBPβ

Expression of *Cebpb* is low in the virgin gland, increases during pregnancy, declines during lactation, and increases again 24 to 48 hours after the onset of involution (Gigliotti and DeWille, 1998; Robinson et al., 1998; Sabatakos et al., 1998). Because a single transcript can give rise to three protein isoforms, the expression and LAP[1] and LIP was assessed (Raught et al., 1995; Dearth et al., 2001). LAP is expressed throughout mammary gland development, with a modest (3-fold) increase during pregnancy, and a small decrease during lactation (Raught et al., 1995; Gigliotti and DeWille, 1998; Sabatakos et al., 1998; Dearth et al., 2001). LIP, historically considered as the dominant negative isoform of C/EBPβ, is present at low levels throughout development, except during pregnancy, when a greater than 100-fold increase in expression is observed (Raught et al., 1995). The robust increase in LIP expression during pregnancy coincides with the proliferation and differentiation of alveolar cells. High levels of LIP are also observed in mouse mammary tumours and in human ductal carcinomas, whereas the normal tissue that surrounds these tumours contains almost undetectable levels of LIP (Raught et al., 1996; Zahnow et al., 1997). The expression of LIP appears to be controlled at the translational level by EGF receptor (EGFR) signalling (Baldwin et al., 2004). EGFR leads to the phosphorylation of an RNA-binding protein called CUG binding protein 1 (CUG-BP1), which in turn binds to the *Cebpb* transcript, and increases the selective

[1] A distinction was not made between LAP* and LAP in these studies; therefore, LAP refers to both activating isoforms.

expression of the LIP isoform (Baldwin et al., 2004). The differential expression of LAP and LIP led several groups to posit that the LAP/LIP ratio may be important for mammary gland development, and has since been shown to play a role in the regulation of the expression of milk protein genes during pregnancy and lactation (Raught et al., 1995; Robinson et al., 1998; Seagroves et al., 1998; Dearth et al., 2001). The LAP/LIP ratio is low (<5) during early pregnancy, when low milk protein gene expression is observed. At parturition, the LAP/LIP ratio increases more than 100-fold, and this increase is correlated with the up-regulation of milk protein genes, such as *Csn2* and *Wap* (Robinson et al., 1998; Seagroves et al., 1998).

C/EBPδ

The expression of *Cebpd* is low in the virgin mammary gland, increases during pregnancy, and decreases again during lactation (Clarkson et al., 2004). *Cebpd* expression increases by more than 100-fold at the onset of involution (Gigliotti and DeWille, 1998; Clarkson et al., 2004), and this rise precedes the expression of pro-apoptotic genes, such as testosterone repressed prostate message 2 [*Trpm2*, (Guenette et al., 1994)], interleukin-1β converting enzyme [*Casp1*, (Boudreau et al., 1995)] and tissue transglutaminase (Strange et al., 1992), as well as genes encoding tissue remodeling proteins, such as metalloproteinases and their inhibitors (Talhouk et al., 1992; Li et al., 1994). These data suggest that C/EBPδ may be involved in the regulation of early apoptotic genes that are required for mammary gland involution. Later studies confirmed that *Cebpd* is directly upregulated by STAT3 at the onset of involution (Thangaraju et al., 2005). C/EBPδ then induces the expression of pro-apoptotic genes, such as *p53*, *Bak* and *Igfbp5*, while concomitantly repressing anti-apoptotic genes, such as *Blf1* and *Ccnd1* (Thangaraju et al., 2005).

1.3.4 C/EBPβ in the mouse mammary gland

General phenotype of three C/EBPβ null mice

The first two C/EBPβ knockout mice were generated by Kishimoto's group and Poli's group to assess the role of C/EBPβ in macrophage-dependent bactericidal and tumoricidal processes (Tanaka et al., 1995) and in hematopoiesis (Screpanti et al., 1995). Both groups generated the *Cebpb* null allele by substituting the bZIP domain with an MC1-Neo poly(A)$^+$ cassette, which inactivates all three isoforms of C/EBPβ (Screpanti et al., 1995; Tanaka et al., 1995). The vector containing the mutant *Cebpb* gene was then used to mutate the endogenous *Cebpb* gene in mouse embryonic stem (ES) cells through homologous recombination. Clones which had undergone correct targeting were injected into recipient blastocysts, and the resultant mutant mice were screened for germline transmission of the mutant allele (Screpanti et al., 1995; Tanaka et al., 1995). The third *Cebpb*-/- mouse was generated by Johnson's group, who replaced a 1.2 kb genomic fragment containing the entire *Cebpb* coding region and some additional promoter sequences with a pGKneobpA cassette (Sterneck et al., 1997). Correctly recombined ES cells clones were injected into blastocysts, and the resultant mutant mice were screened for germline transmission of the mutant allele (Sterneck et al., 1997). *Cebpb*-/- mice were viable, however, homozygous mutants were born at a lower-than-expected Mendelian ratio (Sterneck et al., 1997). Overt defects included sterility in females, splenomegaly, peripheral lymph node enlargement, high susceptibility to infections from microorganisms such as *Candida albicans* and *Listeria monocytogenes*, and a reduction in brown adipose tissue (Screpanti et al., 1995; Tanaka et al., 1995; Darlington et al., 1998). A subset of *Cebpb*-/- mice also exhibited profound hypoglycemia and died shortly after birth (Croniger et al., 1997).

The mammary gland in C/EBPβ null mice

In the mammary gland, *Cebpb*-/- mice exhibit defects in ductal morphogenesis and alveolar development similar to those in the PR, PRLR, GATA3 and STAT5

knockout mice (Lydon et al., 1995; Liu et al., 1997; Ormandy et al., 1997b; Robinson et al., 1998; Seagroves et al., 1998). Mammary gland development is not affected in prepubescent females, with a comparable number of TEBs present at 5 weeks of age in the wild type and mutant mice, albeit a small delay in ductal growth is observed in the mutant mice (Robinson et al., 1998; Seagroves et al., 1998). However, by 8-12 weeks, the ducts become abnormally distended, appearing 'bloated', and fewer side-branches are present when compared to wild type littermates (Robinson et al., 1998; Seagroves et al., 1998). The absence of ovarian hormones in *Cebpb*-/- mice may contribute to the mammary developmental defects, therefore to address this, Robinson transplanted wild type ovaries into *Cebpb*-/- females (Robinson et al., 1998). Ductal morphogenesis was partially restored following ovarian transplants, suggesting that C/EBPβ mediates some of its developmental effects on the mammary gland through indirect mechanisms (Robinson et al., 1998). To demonstrate that the defects observed in ductal development are, at least in part, intrinsic to the mammary epithelium, *Cebpb*-/-, *Cebpb*+/- and wild type mammary glands were transplanted into 3-week old recipients, and the outgrowths were analyzed 6 weeks after transplantation (Robinson et al., 1998; Seagroves et al., 1998). The outgrowths from the *Cebpb*-/- mammary epithelium exhibited defects in ductal branching, similar to the phenotype *in vivo*, whereas the wild type glands were normal. The *Cebpb*+/- outgrowths had an intermediate phenotype, with some ductal distension and reduced branching. To investigate whether C/EBPβ also affects alveolar development, *Cebpb*-/- females were either treated with oestrogen and progesterone to mimic pregnancy (Seagroves et al., 1998), or were mated after receiving a wild type ovary (Robinson et al., 1998). In both experiments, the mammary glands of *Cebpb*-/- mice displayed significant impairment in alveolar development and pregnancy-induced ductal side-branching. *Cebpb*-/- mice developed fewer alveoli than wild type littermates, although the alveoli that were present appeared histologically normal, and displayed a significant reduction in the number of side-branches. Robinson also assessed the role of C/EBPβ in the stroma by co-transplanting wild type mammary epithelium and *Cebpb*-/- stroma into wild type recipients (Robinson et al., 1998). The mammary outgrowths developed normally, including normal ductal branching during puberty, and normal

alveolar development during pregnancy. These data demonstrate that the absence of C/EBPβ in the stroma is not essential for mammary gland development.

C/EBPβ target genes in the mammary gland

To date, no direct *in vivo* target genes have been identified for C/EBPβ in the mammary gland. Through *in vitro* studies, however, the mammary genes *Id2*, *Csn2*, *Xdh* and *Prlr* were shown to be directly regulated by C/EBPβ. *Id2* is a negative regulator of basic helix-loop-helix (bHLH) proteins. Deletion of *Id2* leads to impaired alveolar development due to defects in both the proliferation and survival of alveolar cells during pregnancy (Mori et al., 2000). In *Cebpb*-/- mice, *Id2* expression is reduced, whereas over-expression of *Cebpb* in NIH3T3 cells leads to up-regulation of *Id2* expression (Karaya et al., 2005). These data suggested that *Id2* may be a direct C/EBPβ target gene, and this was subsequently confirmed through ChIP experiments in NIH3T3 cells, where C/EBPβ binding was observed in the *Id2* proximal promoter (Karaya et al., 2005). *Csn2* and *Xdh* are prominent components of milk, and their expression is directly regulated by C/EBPβ in HC11 cells (a mammary epithelial cell line) (Kim et al., 2002; Seymour et al., 2006). More recently, *Prlr* was also identified as a direct C/EBPβ target gene in a human breast cancer cell line (T47D) (Goldhar et al., 2011).

2 Two lineage-primed luminal progenitor cell populations exist in the mouse mammary gland

2.1 Introduction

Many questions remain regarding the organization of the cellular hierarchy in the mouse mammary gland. Stingl, Shackleton and Asselin-Labat identified a luminal progenitor population (Ma-CFC) using slightly different cell surface markers: $Lin^{neg}CD24^{hi}CD49f^{lo}$, $Lin^{neg}CD24^{hi}CD29^{lo}$, $Lin^{neg}CD24^{hi}CD29^{lo}CD61^{pos}$, respectively (Asselin-Labat et al., 2006; Shackleton et al., 2006; Stingl et al., 2006). These cells express luminal keratins (e.g., K8, K18) and form smooth, round colonies when cultured *in vitro*. Shackleton also showed that Ma-CFCs produce milk proteins when cultured under lactogenic conditions (Shackleton et al., 2006). Together, these studies demonstrated that the adult mammary gland contains luminal-lineage specific progenitor cells. However, they did not address how the ductal and alveolar lineages develop from within this population. It therefore remains unclear whether the two luminal cell types (ductal and alveolar cells) are maintained by one bipotent luminal progenitor population, or by two unipotent progenitor populations within the Ma-CFC.

Sleeman demonstrated that $CD45^{neg}CD24^{hi}$ luminal cells can be divided into two populations based on Prom1 (CD133) expression, and that CD133 and Sca1 expression are concordant in luminal epithelial cells (Sleeman et al., 2007). $CD133^{pos}$ luminal cells express five hormone-sensing genes: *Esr1*, *Prlr*, *Pgr*, *Cited1* and *S100a6*, whereas $CD133^{neg}$ luminal cells express three milk protein genes: *Csn2*, *Wap* and *Ltf*. Both populations have poor repopulating capacity *in vivo*, but high clonogenic activity *in vitro*. These data suggest that the $CD133^{pos}$ and $CD133^{neg}$ luminal epithelial cell populations contain precursors to the ductal and alveolar lineages, respectively. However, a subsequent study by Kendrick found higher expression of some alveolar genes (e.g., *Csn2*, *Csn1s2a*) in the $Sca1^{pos}$ population

(Kendrick et al., 2008). Therefore, it is presently unclear whether the Sca1^{pos} and Sca1^{neg} luminal progenitor populations are lineage-biased.

2.2 Aims of this chapter

The first aim of this chapter is to describe the development and optimization of a staining strategy that enabled a more precise isolation of Sca1^{pos} and Sca1^{neg} luminal progenitor cells. First, the relationship between the different progenitor populations published in the studies mentioned previously was evaluated by direct comparison of the gating strategies. To achieve this, cells from dissociated adult (10-12 weeks old) virgin mammary glands were stained with a combination of the antibodies used in the above publications, and the different gating strategies were compared directly. Based on these analyses, we developed a new staining strategy that incorporated many of the cell surface markers used in the previous studies.

The second aim of this chapter is to describe the gene expression profile and *in vitro* differentiation potential of the two luminal progenitor cell populations we identified. The populations we isolated are similar to Sleeman and Kendrick's (Sleeman et al., 2007; Kendrick et al., 2008), as they differentially express Sca1. To clarify the difference they observed in alveolar gene expression in the Sca1^{pos} luminal cells, as well as to determine whether Sca1^{pos} and Sca1^{neg} luminal cells are biased in their gene expression towards ductal and alveolar cell fates, respectively, we performed quantitative reverse transcriptase polymerase chain reaction (qRT-PCR) and microarray analysis. We expected that if differential lineage priming is observed in the two luminal progenitors, then upon *in vitro* differentiation, each would generate colonies with a gene expression pattern and morphology that reflects their lineage priming. We therefore performed *in vitro* colony forming assays and assessed the gene expression patterns of the colonies generated by Sca1^{pos} and Sca1^{neg} luminal cells using qRT-PCR.

2.3 Collaborators

The work described in this chapter was performed in collaboration with by Dr. Susana Garcia-Silva. Dr. Silva developed the fluorescence-activated cell sorting (FACS) staining strategy, prepared virgin luminal cells for microarray, and performed qRT-PCR and colony assays. Figures 2.7-2.9 were generated from Dr. Silva's experiments. I helped Dr. Silva develop the mammary gland staining strategy, performed qRT-PCR and colony assays, analyzed the microarray data, and generated all of the figures. Figures 2.1-2.6, and 2.10 were generated from my own experiments. Dr. Simon Tomlinson and Susan Moore curated the microarray data.

2.4 Results

2.4.1 Comparison of flow cytometric strategies used to isolate luminal progenitors

To directly compare the different staining strategies used in the studies described in the introduction, flow cytometric analysis of fresh mammary tissue isolated from 10-week old virgin wild type females was performed. The published studies performed on adult virgin females (age 10-12 weeks old) used different combinations of lineage markers, and these differences may affect the direct comparison of the gating strategies. Therefore, a common strategy was used in these analyses, whereby non-singlets, dead cells (7-aminoactinomycinD [7-AAD]^{pos}) and lineage positive cells were first excluded. Lineage is defined in this thesis by the markers CD45, CD31 and Ter119. Inclusion of the lineage marker CD140a was tested by Dr. Mario Buono, who noted that the frequency of luminal and myoepithelial cells did not change upon addition of this marker (data not shown). The fluorescence minus one (FMO) controls used to set the gates are shown in Appendix 1.

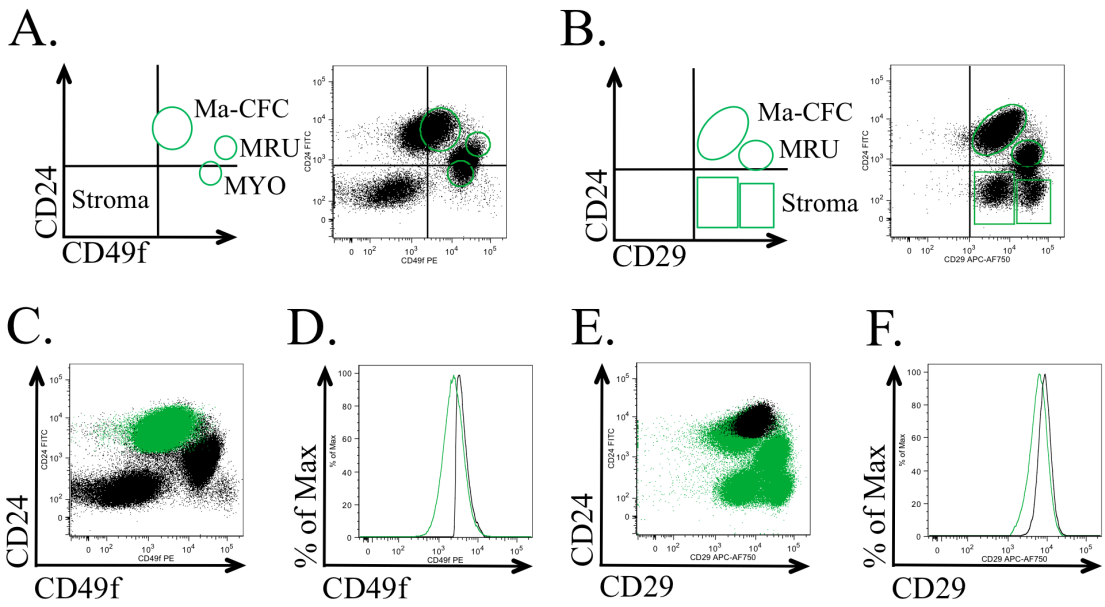


Figure 2.1 Comparison between the flow cytomteric strategies of Stingl and Shackleton. (A&B) Stingl's staining strategy (A) and Shackleton's staining strategy (B). Dot plots were first gated on viable singlets and Lin (CD45, CD31, Ter119) cells (not shown). The four quadrants created with black lines were set using fluorescence minus one controls to indicate negative & positive positions within the graph **(C&D)** Shackleton's Ma-CFCs (green) analyzed in Stingl's gating strategy (black). **(E&F)** Stingl's Ma-CFC (black) analyzed in Shackleton's gating strategy (green). These plots show that the Ma-CFC population identified by the two staining strategies marks the same population of cells. **Abbreviations:** Ma-CFC, mammary colony forming cells; MRU, mammary repopulating unit; MYO, myoepithelial cells; Max, maximum.

Comparison between Stingl and Shackleton's staining strategy is shown in Figure 2.1 (Shackleton et al., 2006; Stingl et al., 2006). Stingl's staining divides the Lin^{neg} compartment into three populations based on CD24 and CD49f expression (Fig. 2.1A). The luminal compartment (Ma-CFC) is defined as $CD24^{hi}CD49f^{lo}$ and forms 23 % of the Lin^{neg} gate. Shackleton's staining divides the Lin^{neg} compartment into four populations based on CD24 and CD29 expression (Fig. 2.1B). The luminal compartment (Ma-CFC) is defined as $CD24^{hi}CD29^{lo}$ and forms 58 % of the Lin^{neg} gate. Stingl's Ma-CFC population (black) assessed for CD29 expression includes all of Shackleton's Ma-CFCs (green), albeit Stingl's Ma-CFCs are restricted to the $CD29^{med}$ population, and do not include the $CD29^{lo}$ cells (Fig. 2.1C,D). Conversely, Shackleton's Ma-CFCs (green) assessed for CD49f expression include all of Stingl's Ma-CFCs (black), although Shackleton's gating also includes some CD49f negative/low cells not included in Stingl's gating strategy (Fig. 2.1E,F). These cells form a part of the same cell population, therefore this discrepancy is due to the stringency in the cutoff for CD49f expression used by Stingl. These analyses show that the Ma-CFCs investigated by Stingl and Shackleton are overlapping populations, and have the combined cell surface phenotype of $Lin^{neg}CD24^{hi}CD49f^{lo}CD29^{lo}$.

Asselin-Labat's gating strategy was then compared with Stingl and Shackleton's (Fig. 2.2) (Asselin-Labat et al., 2006). Asselin-Labat first gated on $Lin^{neg}CD24^{pos}$ cells, and then compared the expression of CD61 and CD29 (Fig. 2.2A). Both the $CD29^{lo}CD61^{pos}$ and $CD29^{lo}CD61^{neg}$ populations (green) co-localize with Stingl's (Fig. 2.2B,C) and Shackleton's (Fig. 2.2D,E) Ma-CFCs (black).

Sleeman used the cell surface markers CD24, CD133 and Sca1 to distinguish between three cell populations (Fig. 2.3) (Sleeman et al., 2007). The $CD24^{hi}$ cells can be divided into two cell populations based on CD133 expression. However, Sleeman also noted that Sca1 and CD133 expression are concordant. Therefore, Sca1 was substituted for CD133 in these analyses (Fig. 2.3A). When compared with Stingl's (Fig. 2.3B,C) and Shackleton's (Fig. 2.3D,E) staining, both the $CD24^{hi}Sca1^{pos}$ (upper panel in Fig. 2.3B-E) and $CD24^{hi}Sca1^{neg}$ cells (lower panel in Fig. 2.3B-E) co-localize with the Ma-CFCs (black).

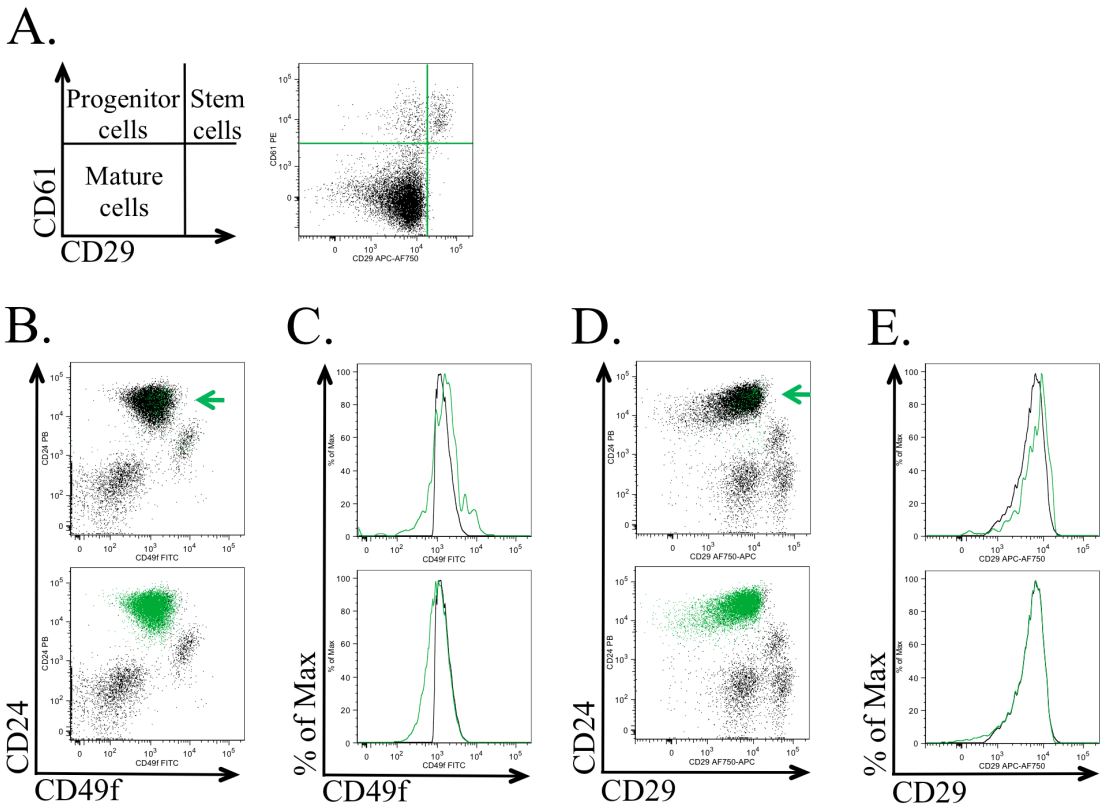


Figure 2.2 Comparison between the flow cytometric strategy of Asselin-Labat, Stingl and Shackleton. (A) Asselin-Labat's gating strategy. The dot plot was first gated on viable singlets, lineage (CD45, CD31, Ter119) negative cells and CD24 positive cells (not shown) prior to evaluating CD61 and CD29 expression. The two $CD29^{lo}$ quadrants (progenitor cells-top panels in B-E, and mature cells-bottom panels in B-E) from (A) were then compared to Stingl's (**B&C** in black) and Shackleton's (**D&E** in black) staining. Both compartments (in green) co-localize with the mammary colony forming cell population. **Abbreviations:** Max, maximum.

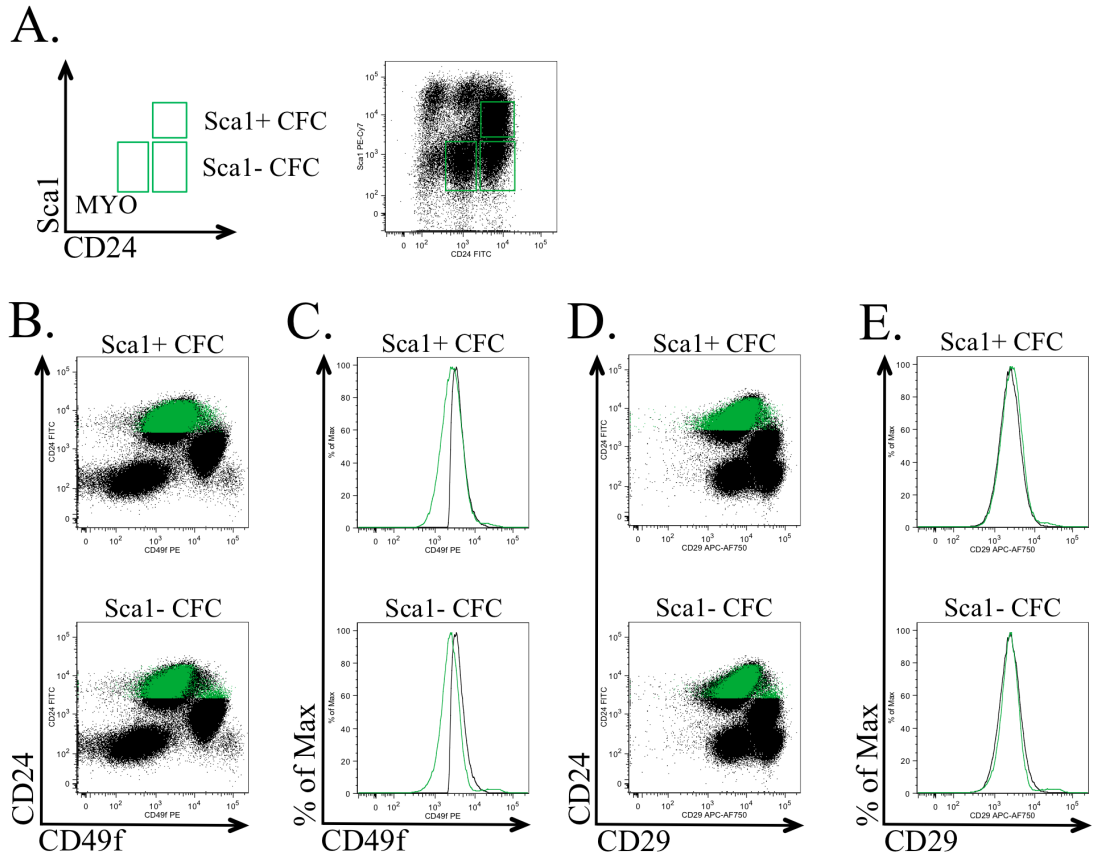


Figure 2.3 Comparison between the flow cytometric strategy of Sleeman, Stingl and Shackleton. (A) Sleeman's gating strategy. The original gating used by Sleeman was based on CD133 and CD24 expression. However, he also noted that Sca1 and CD133 expression are concordant, therefore, Sca1 was substituted for CD133 in these analyses. The dot plot in (A) was first gated on viable singlets and lineage (CD45, CD31, Ter119) negative cells (not shown) prior to evaluating Sca1 and CD24 expression. $Sca1^{pos}$ (top panel in B-E, green) and $Sca1^{neg}$ (bottom panel in B-E, green) were compared with Stingl's (**B&C**) and Shackleton's (**D&E**) mammary colony forming cells. Both cell populations co-localize with Ma-CFCs, and a portion also co-localize with MRUs ($CD49f^{hi}$ in B and $CD29^{hi}$ in D). **Abbreviations:** MYO, myoepithelial cells; CFC, colony forming cells; Ma-CFC, mammary colony forming cells; Max, maximum.

A small portion of $CD24^{hi}Sca1^{neg}$ cells also appear to mark a subset of myoepithelial cells, which express intermediate levels of CD24 and CD29.

Lastly, Li first gated on $CD24^{hi}$ cells, and then assessed Sca1 and CD49b expression (Fig. 2.4) (Li et al., 2009). Li observed that the two $CD49b^{pos}$ populations ($Sca1^{pos}$ and $Sca1^{neg}$) are enriched in progenitor activity when compared to the $CD49b^{neg}$ cells. All three populations co-localize with Ma-CFCs (Fig. 2.4B-E). However, similar to Sleeman's staining, some of the $Sca1^{neg}$ luminal cells also contain a subset of cells from the myoepithelial compartment.

To achieve the highest possible purity of $Sca1^{pos}$ and $Sca1^{neg}$ luminal cells, we developed a staining strategy that combined the cell surface markers employed by the above research groups (Fig. 2.5). First, non-singlets, dead cells and lineage positive cells were excluded. Next, cells were gated based on CD29 expression ($CD29^{pos}$), which included two distinct cell populations: $CD29^{lo}$ and $CD29^{hi}$. The $CD29^{lo}$ population co-localizes with the Ma-CFCs described by Stingl and Shackleton, and the $CD29^{hi}$ population corresponds to the myoepithelial/MRU (mammary repopulating unit or stem cell) population. Next, we assessed CD24 and CD49f expression, where we obtain the same three groups of cells as Stingl (Fig. 2.5B,D). In our staining, the Ma-CFCs (marked in solid green color in Fig. 2.5A, 5th panel from the left) make up 57 % of the Lin^{neg} population, similar to Shackleton's gating (58 %) (Fig. 2.5C,E). Using these gates in addition to the assessment of Sca1 and CD49b expression in the luminal compartment allowed us to remove the small subset of myoepithelial cells that were included in Sleeman's and Li's isolation strategy (Fig. 2.3 & Fig. 2.4). These myoepithelial cells would fall into the $Sca1^{neg}$ gate in the subsequent plot (Appendix 1). In the representative plot in Figure 2.5, the $Sca1^{pos}$ and $Sca1^{neg}$ luminal cells are 27 % and 21 % of the Lin^{neg} population, respectively. In Sleeman's gating, the $Sca1^{pos}$ cells are 27 % and the $Sca1^{neg}$ cells are 25 % of the Lin^{neg} population. Of note is that the frequency of the two populations varies between 10-week old, wild type virgin females from 20-32 % of Lin^{neg}, with an average of 24.55 ± 5.11 % (mean ± standard deviation (SD); n=4) for $Sca1^{pos}$ cells, and 27.65 ± 1.64 % (n=4) for $Sca1^{neg}$ cells. To verify that the two cell

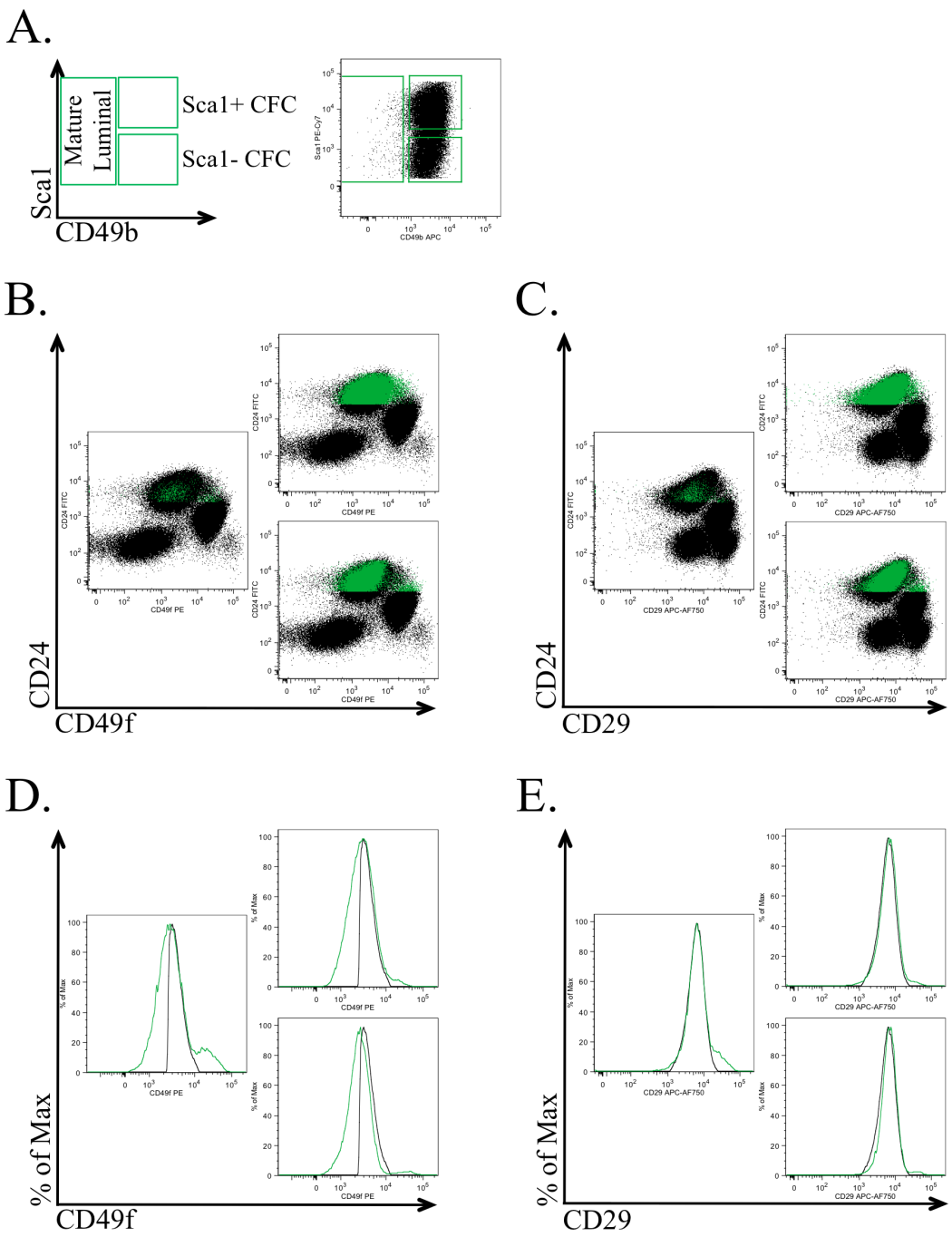


Figure 2.4 Comparison between the flow cytometric strategy of Li, Stingl and Shackleton. A) Li's gating strategy. The dot plot was first gated on viable singlets, lineage (CD45, CD31, Ter119) negative cells and CD24 high cells (not shown) prior to evaluating Sca1 and CD49b expression. The three panels in (A) were compared to Stingl's (B&D) and Shackleton's (C&E) gating strategies, and shows that all three populations co-localize with the mammary colony-forming cell population. **Abbreviations:** CFC, colony-forming cells; Max, maximum.

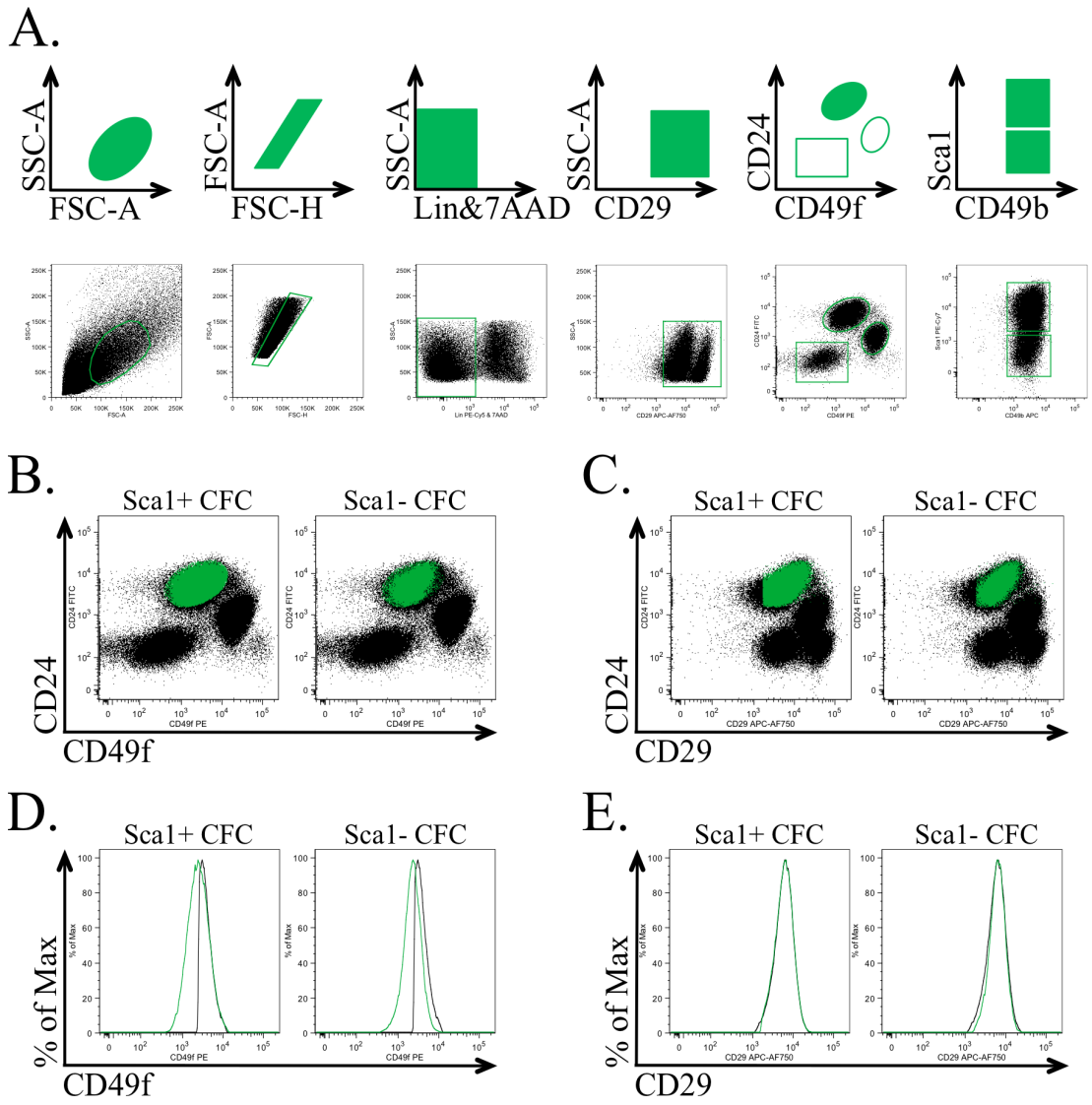
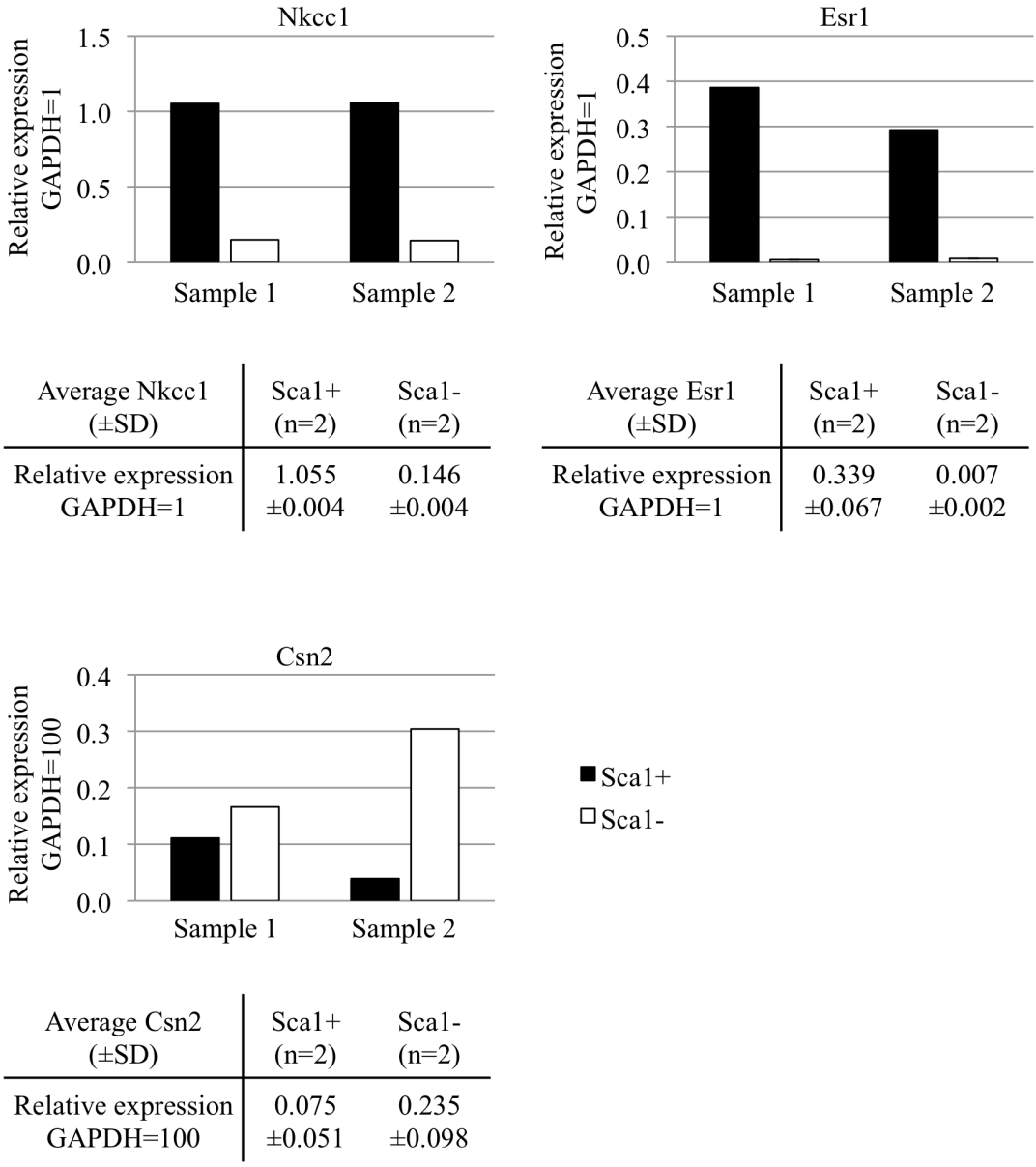


Figure 2.5 Comparison between the flow cytometric strategy developed by Garcia-Silva and Zay (this thesis), Stingl and Shackleton.
(A) Schematic diagram (top row) and representative dot plots (bottom row) depicting the gating strategy used to isolate Sca1^{pos} and Sca1^{neg} luminal progenitor cells from adult (10-12 week old) virgin mice. Cells were first gated on scatter to obtain singlets. Then lineage (CD31, CD45, Ter119) positive cells were excluded, and Lin^{neg} cells were selected based on CD29 expression (CD29lo correspond to Ma-CFCs and CD29hi cells correspond to MYO/MRUs). CD29^{pos} cells were then evaluated for CD24 and CD49f expression. CD24^{hi}CD49f^{lo} cells were further subdivided based on Sca1 expression, and both Sca1^{pos} and Sca1^{neg} luminal cells co-localize with the mammary colony forming cells identified by Stingl **(B&D)** and Shackleton **(C&E)**. **Abbreviations:** SSC-A, side scatter area; FSC-A, forward scatter area; FSC-H, forward scatter height; CFC, colony forming cell; Max, maximum.



Average Nkcc1 (±SD)	Sca1+ (n=2)	Sca1- (n=2)
Relative expression GAPDH=1	1.055 ±0.004	0.146 ±0.004

Average Esr1 (±SD)	Sca1+ (n=2)	Sca1- (n=2)
Relative expression GAPDH=1	0.339 ±0.067	0.007 ±0.002

Average Csn2 (±SD)	Sca1+ (n=2)	Sca1- (n=2)
Relative expression GAPDH=100	0.075 ±0.051	0.235 ±0.098

Figure 2.6 Relative expression of selected ductal and alveolar genes in Sca1^{pos} and Sca1^{neg} luminal cells. Quantitative real-time PCR analysis of ductal (*Nkcc1* & *Esr1*) and alveolar (*Csn2*) genes was performed on sorted Sca1^{pos} (black bars) and Sca1^{neg} (white bars) luminal cells from 12-week old wild type virgin females. The values represent gene expression normalized to *Gapdh* from two biological replicates (Sample 1 and Sample 2), each assayed in technical triplicate. The table below each graph shows the average value from the two replicates ± standard deviation. **Abbreviations:** SD, standard deviation; n, number.

populations we sorted were comparable to Sleeman's, qRT-PCR was performed on *Esr1*, *Csn2*, and *Nkcc1* (Figure 2.6). The ductal genes *Nkcc1* and *Esr1* were highly expressed in Sca1^{pos} cells when compared to Sca1^{neg} cells. The alveolar gene *Csn2* was more highly expressed in Sca1^{neg} cells, albeit not significantly. The overall expression of *Csn2* was approximately 100-fold lower than *Nkcc1* and 10-fold lower than *Esr1*.

2.4.2 The luminal compartment contains two differentially programmed progenitor populations

To address how ductal and alveolar specification might occur in the luminal progenitor compartment, the Sca1^{pos} and Sca1^{neg} luminal cells were assessed for gene expression by microarray and qRT-PCR. Gene profiling using Affymetrix exon arrays identified 936 differentially expressed genes (false detection rate [FDR] $\leq$ 0.01) between the Sca1^{pos} (776 genes) and Sca1^{neg} cells (160 genes) (Fig 2.7A). The genes with higher expression in Sca1^{pos} luminal cells included the previously reported hormone sensing genes *Prlr*, *Pgr* and *Esr1*, as well as other genes associated with the ductal epithelium, for example *Areg* and *Foxa1* (Fig. 2.7B). Genes that were preferentially expressed in the Sca1^{neg} luminal cells included milk protein genes, such as *Xdh*, *Ltf* and *Csn3*, and the gene encoding the alveolar-specific transcription factor *Elf5* (Fig 2.7B). The calcium binding protein gene *S100a6* and the milk protein gene *Csn1s2a* were not differentially expressed between the two populations (data not shown). To confirm that the ductal (D-genes) and alveolar (A-genes) genetic programs were enriched in the Sca1^{pos} and Sca1^{neg} luminal cells, respectively, microfluidics-based real time PCR was performed. The expression of the five selected ductal genes (*Prlr*, *Pgr*, *Areg*, *Esr1*, *Foxa1*) was enriched in the Sca1^{pos} cells, whereas the four alveolar genes (*Csn3*, *Xdh*, *Ltf*, *Elf5*) were enriched in the Sca1^{neg} cells (Fig. 2.7C). Common luminal genes (L-genes), such as *Krt18*, *Sox9*, *Klf5* and *Gata3*, were highly but not differentially expressed in the two populations (Fig. 2.7B-C). Functional annotation clustering analysis using the DAVID bioinformatics resource (Anderson et al., 2007; Huang et al., 2008) showed that the overall D-gene signature is enriched in genes associated with hormone response,

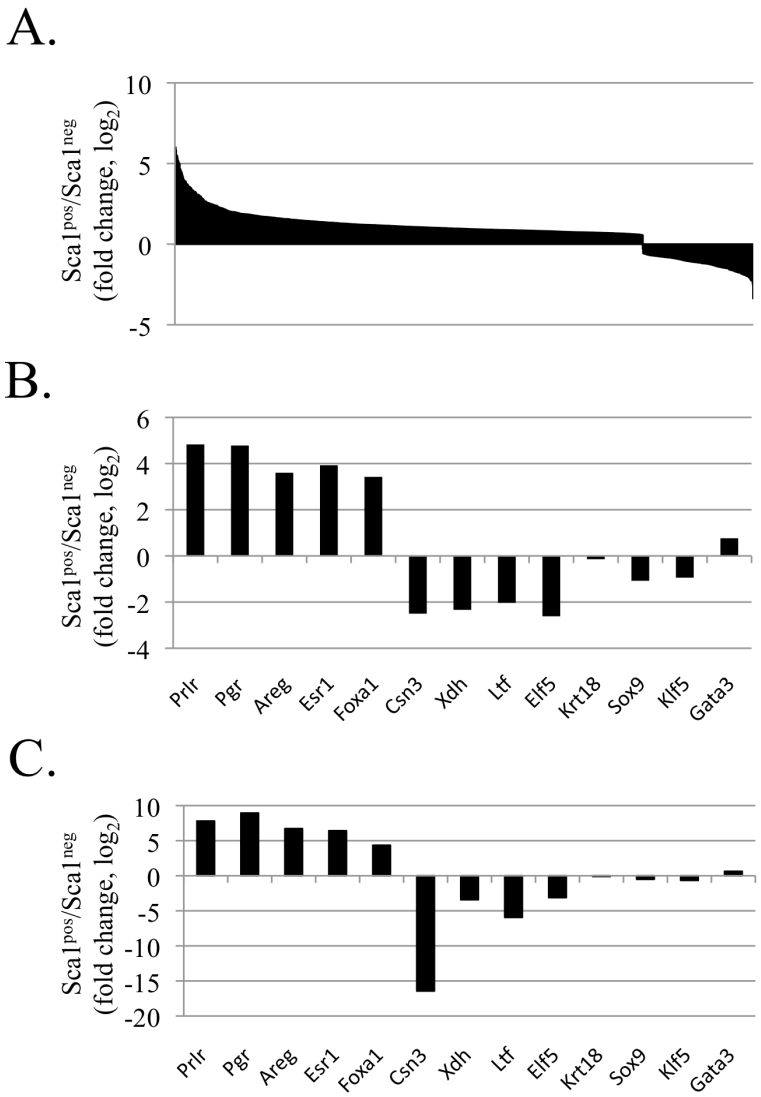
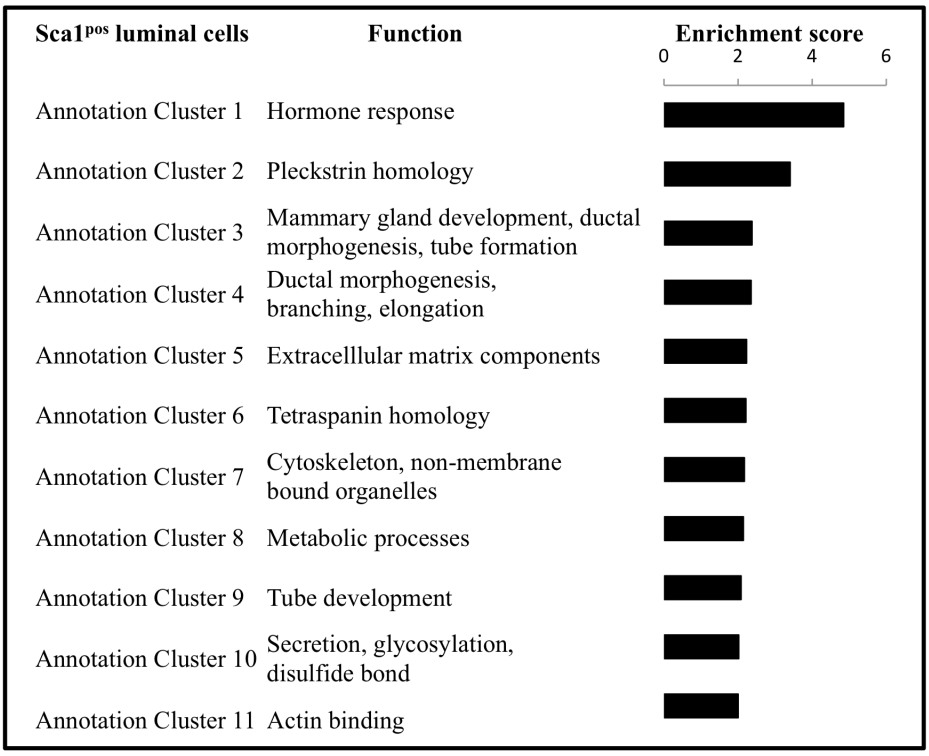


Figure 2.7 Differentially expressed genes in Sca1^{pos} and Sca1^{neg} luminal cells. (A) Differentially expressed genes between Sca1^{pos} and Sca1^{neg} luminal cells that were identified from the microarray analysis (FDR ≤0.01) are represented as the ratio of Sca1^{pos} to Sca1^{neg} values. The RMA-normalized, annotated, and quality-filtered expression values were tested for differential expression using Limma in the BioConductor framework, and the values shown are the ratio between the average Sca1^{pos} cells (n=3) and Sca1^{neg} cells (n=3) expression values after $\log_2$ transformation. **(B)** Selected genes from (A) showing that hormone sensing genes (*Prlr*, *Pgr*, *Esr1*) are enriched in Sca1^{pos} luminal cells, whereas milk genes (*Csn3*, *Xdh*) are enriched in Sca1^{neg} luminal cells. Highly but not differentially expressed genes are shown for comparison (*Krt18*, *Sox9*, *Klf5*, *Gata3*). (**C)** Quantitative real-time PCR validation of microarray data. Gene expression was determined by quantitative real-time PCR on Fluidigm 48.48 dynamic arrays. For each gene, 4 biological replicates were analyzed in triplicate for each population, and the value shown is the ratio of the average Sca1^{pos} and Sca1^{neg} values after $\log_2$ transformation.

mammary gland morphogenesis, branching and tube formation (Fig. 2.8A). Whereas the overall A-gene signature is enriched in genes associated with secretion, cell motility, extracellular matrix interactions and ion transport (Fig. 2.8B). Comparison between the small population of Sca1^{pos}CD49b^{neg} and the Sca1^{pos}CD49b^{pos} cells identified only 19 differentially expressed genes (FDR $\leq$ 0.01) (not shown), indicating that these two cell populations are the same. Together, these data suggest that the Sca1^{pos} and Sca1^{neg} luminal populations contain functionally distinct progenitors biased towards ductal and alveolar differentiation, respectively.

To determine whether upon *in vitro* differentiation the Sca1^{pos} and Sca1^{neg} luminal cells would generate colonies that reflect their *in vivo* lineage priming with respect to gene expression pattern and colony morphology, freshly sorted cells were cultured in semi-solid medium in the presence of PRL and RANKL (Fig. 2.9), or heparin and EGF (Figure 2.10). PRL and RANKL promote the differentiation of alveolar and ductal cells. In Figure 2.9, Sca1^{pos} and Sca1^{neg} luminal cells were cultured for 15 days, with the medium supplemented with PRL and RANKL for the last 7 days to promote differentiation in the colonies. While both Sca1^{pos} and Sca1^{neg} progenitors formed smooth, round colonies in the absence of PRL/RANKL, only the Sca1^{pos} luminal progenitors formed irregular shaped colonies that resemble branching (Fig. 2.9A). Gene expression analysis of the colonies after PRL/RANKL treatment showed that the A-signature was maintained in the Sca1^{neg} progenitors, whereas the D-signature was maintained in the Sca1^{pos} progenitors (Fig. 2.9B). These data suggest that the Sca1^{pos} and Sca1^{neg} luminal cells exhibit distinct responses to PRL and RANKL, consistent with their *in vivo* transcriptional identity. To further demonstrate that *in vitro* branching morphogenesis was specific to the Sca1^{pos} luminal cells, both Sca1^{pos} and Sca1^{neg} populations were cultured with EGF, FGF and heparin (Fig. 2.10). After 21 days in culture, the Sca1^{pos} cells formed branched colonies, whereas the Sca1^{neg} cells formed smooth, round colonies (Fig. 2.10A,C). The frequency of colony forming cells was approximately 2-fold higher and the frequency of branching was approximately 9-fold higher in Sca1^{pos} cells when compared with the Sca1^{neg} cells (Fig. 2.10B,C).

A.



B.

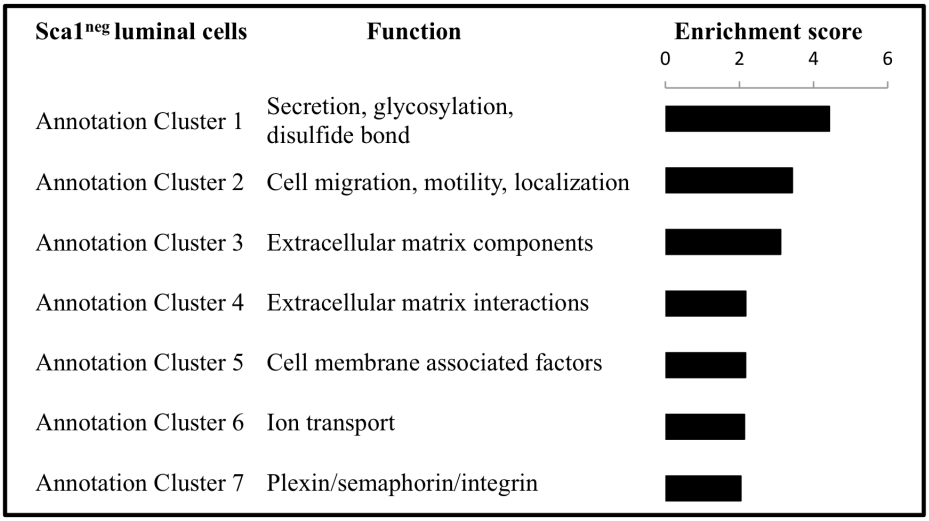


Figure 2.8 Functional annotation clustering analysis of differentially expressed genes in Sca1^{pos} and Sca1^{neg} luminal progenitor cells. (A) Functional clusters generated by enriched genes in Sca1^{pos} luminal progenitors. The transcript cluster IDs of differentially expressed genes (from Fig. 2.7) were analyzed using the DAVID Bioinformatics Resource using default parameters. Functional clusters were included in the figure if they had an enrichment score of ≤2. **(B)** Functional clusters were generated by enriched genes in Sca1^{neg} luminal cells using the same analysis as in (A). Supplementary information related to this figure (complete list of functional clusters) is shown in Appendices 2&3.

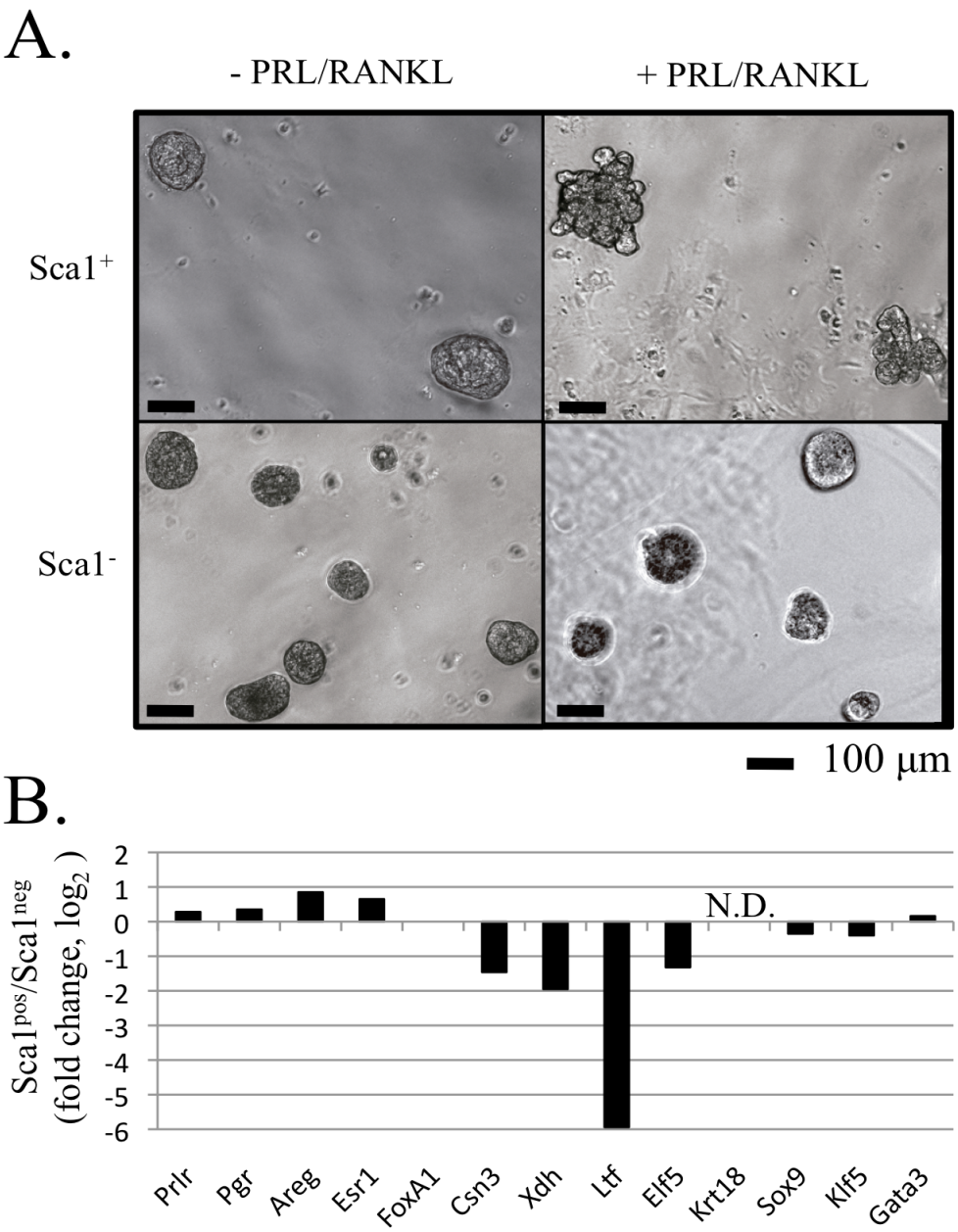


Figure 2.*9 In vitro* colony forming assay of Sca1^{pos} and Sca1^{neg} luminal cells. (A) Representative images of Sca1^{pos} and Sca1^{neg} colonies embedded in Matrigel, and cultured over a layer of irradiated NIH3T3 fibroblast cells for 15 days. On the left, colonies were cultured without PRL/RANKL for 15 days, and on the right, colonies were cultured with PRL/RANKL for the last 7 days of the 15-day culture period. Scale bar=100 µm. **(B)** Gene expression in colonies generated by Sca1^{pos} and Sca1^{neg} luminal cells after 15 days in culture, with the addition of PRL/RANKL for the last 7 days. Gene expression was determined by quantitative real-time PCR on Fluidigm 48.48 dynamic arrays. For each gene, 4 biological replicates were analyzed in triplicate for each population, and the value shown is the ratio of the average Sca1^{pos} and Sca1^{neg} values after log_2 transformation. **Abbreviations:** PRL, prolactin; RANKL, RANK ligand, N.D., not determined.

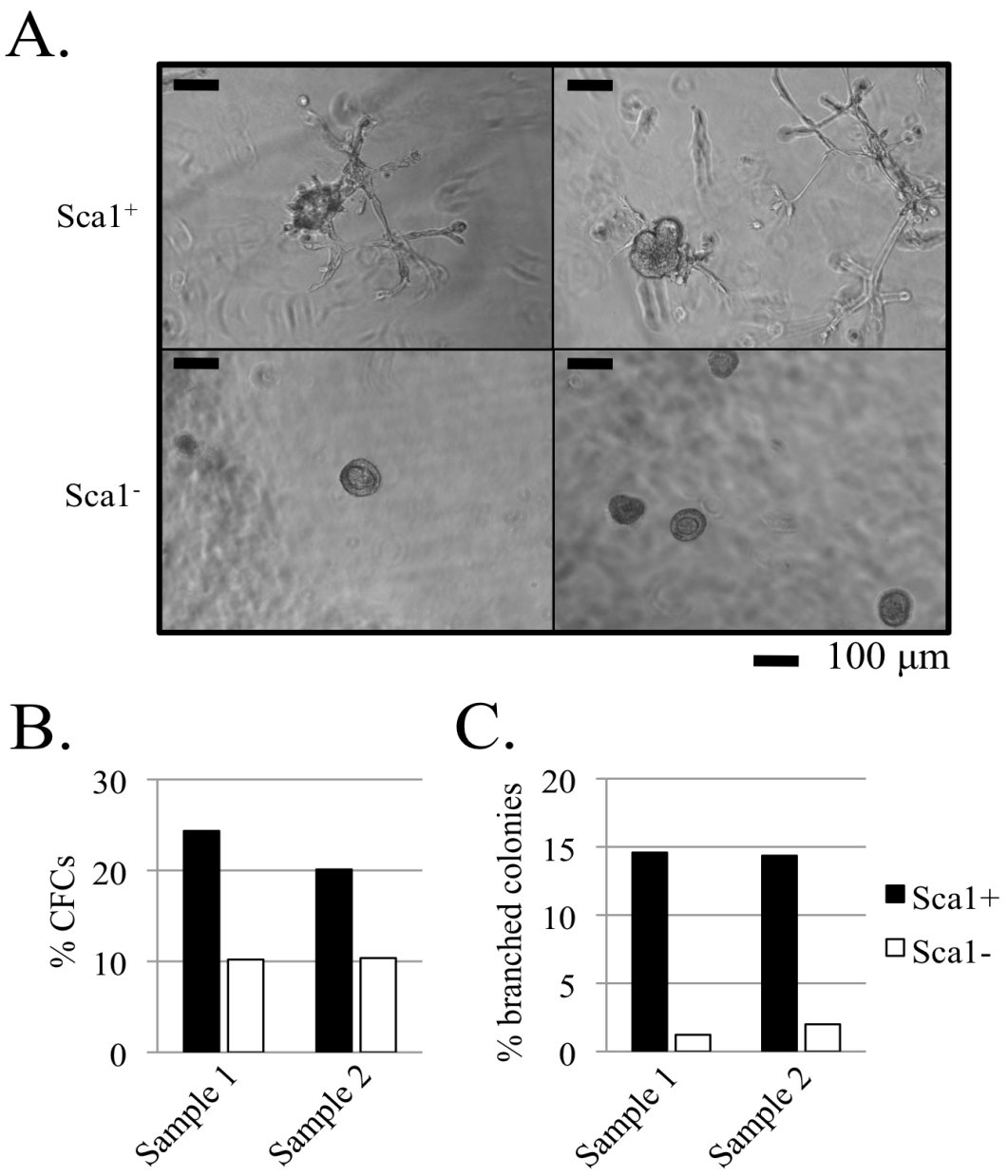


Figure 2.10 *In vitro* colony forming assay of Sca1^{pos} and Sca1^{neg} luminal cells. (A) Representative images of Sca1^{pos} and Sca1^{neg} colonies embedded in Matrigel, and cultured over a layer of mitomycin-C treated NIH3T3 cells in the presence of EGF, FGF and heparin for 21 days. Scale bar=100 µm. **(B)** Quantification of colony forming activity by Sca1^{pos} and Sca1^{neg} luminal cells. Values represent the number of colonies as a frequency of the total number of cell plated from two biological replicates plated in duplicate. **(C)** Quantification of branching in Sca1^{pos} and Sca1^{neg} colonies after 21 days of culture. Values represent the number of branched colonies divided by the total number of colonies. **Abbreviations:** CFC, colony-forming cells.

Together, the gene expression analyses and *in vitro* colony forming assays demonstrate that the Sca1^{pos} luminal compartment contains ductal-primed luminal progenitors (termed ductal luminal progenitors or DLPs), whereas the Sca1^{neg} luminal compartment contains alveolar-primed luminal progenitors (termed alveolar luminal progenitors or ALPs).

2.5 Discussion

2.5.1 Isolation of two luminal progenitor populations

The development of the mammary gland requires the specification of myoepithelial, ductal and alveolar lineages from mammary stem and progenitor cells. Others have previously shown that the luminal epithelial compartment contains progenitors (Ma-CFCs) that express luminal keratins and have high clonogenic activity *in vitro* (Asselin-Labat et al., 2006; Shackleton et al., 2006; Stingl et al., 2006). Furthermore, it has been demonstrated that the luminal cells can be divided into a hormone-sensing population (CD24^{hi}CD133^{pos}) that co-expresses the prolactin, progesterone and oestrogen receptors, and a CD24^{hi}CD133^{neg} population that expresses the alveolar-specific lactoferrin (*Ltf*), β-casein (*Csn2*) and whey acidic protein (*Wap*) genes (Sleeman et al., 2006; 2007). However, CD133 and Sca1 are concordantly expressed in luminal epithelial cells, and a subsequent study by Kendrick found that some alveolar genes (*Csn2*, *Csn1s2a*) were enriched in the CD24^{hi}Sca1^{pos} population, rather than the CD24^{hi}Sca1^{neg} cells (Kendrick et al., 2008). To clarify whether the Sca1^{pos} and Sca1^{neg} luminal cells were lineage-primed towards the ductal and alveolar cell fates, respectively, we characterized their global gene expression and *in vitro* differentiation potential.

The flow cytometric strategy we developed incorporated many of the cell surface markers used by different research groups (Fig. 2.1-2.5). The strategy of gating on Lin, CD29, CD49f and CD24 expression prior to analyzing Sca1 and CD49b expression allowed us to exclude contamination from other compartments, particularly from the myoepithelium, which was observed in our analysis of the

published gating strategies (Fig. 2.3 & 2.4). The initial validation of the sorting strategy was performed by qRT-PCR using three genes with known functions in the mammary gland (Fig. 2.6). *Nkcc1* encodes a sodium-potassium-calcium co-transporter that was shown to localize exclusively to the basolateral membrane of ductal luminal cells by immunofluorescence staining (Shillingford et al., 2002a). *Esr1* expression was observed only in $CD24^{hi}CD133^{pos}$ luminal cells, and immunofluorescence staining demonstrated nuclear ESR1 only in the ductal epithelium (Sleeman et al., 2007). Lastly, *Csn2* encodes a milk protein that is expressed during pregnancy and lactation by alveolar cells (Moreira et al., 2010). The qRT-PCR results from sorted $Sca1^{pos}$ and $Sca1^{neg}$ luminal cells show that *Nkcc1*and *Esr1* are highly expressed in the $Sca1^{pos}$ cells (Fig. 2.6), consistent with previous observations (Shillingford et al., 2002a; Sleeman et al., 2007), whereas the alveolar gene *Csn2* is enriched in $Sca1^{neg}$ cells. *Csn2* did not pass the FDR filter of ≤ 0.01 in the microarray data from 3 biological replicates, demonstrating that it is not differentially expressed between $Sca1^{pos}$ and $Sca1^{neg}$ cells. Consistent with its biological function as a component of milk, we and others observed that the level of expression for *Csn2* is very low in the virgin gland, with expression increasing during pregnancy and lactation (Shillingford et al., 2002a; Moreira et al., 2010). The variability between biological samples and the low level of expression in the virgin mammary gland are the likely reasons for the opposing observations by Sleeman and Kendrick regarding the relative expression of *Csn2* in $Sca1^{pos}$ and $Sca1^{neg}$ luminal cells (Sleeman et al., 2007; Kendrick et al., 2008). Of the five casein genes included in the microarray (*Csn1s1*, *Csn1s2a*, *Csn1s2b*, *Csn2*, *Csn3*), only *Csn3* (expressed more highly in $Sca1^{neg}$ cells) passed the FDR filter of ≤ 0.01.

2.5.2 The luminal compartment contains two lineage-primed progenitors

Following the qRT-PCR validation of the two sorted luminal cell populations, we performed microarray analysis to investigate their global gene signatures. The exon arrays show that $Sca1^{pos}$ and $Sca1^{neg}$ cells differentially express 936 genes (Fig. 2.7A). As expected, genes more highly expressed in $Sca1^{pos}$ cells include the genes described by Sleeman (*Prlr*, *Pgr*, *Esr1, Cited1*), as well as other genes that are

associated with ductal morphogenesis, such as amphiregulin (*Areg*) and forkhead box protein A1 (*Foxa1*) (Sleeman et al., 2007). The expression profile of these genes was confirmed by qRT-PCR (Fig. 2.7C). AREG is an EGFR ligand that is upregulated in the mammary gland during ductal elongation (LaMarca and Rosen, 2007). Ciarloni demonstrated that *Areg* is directly regulated by ESR1, and is an important mediator of oestrogen-induced ductal morphogenesis in the pubertal mammary gland (Ciarloni et al., 2007). Similar to the mammary gland of *Esr1*-/- females, *Areg*-/- mammary glands have significant developmental defects in TEB formation, ductal elongation and branching during puberty. However, unlike in *Esr1*-/- mice, pregnancy-induced side-branching and alveolar differentiation in *Areg*-/- females are not disrupted. These data demonstrate that the function of AREG is restricted to ductal luminal cells. FOXA1 is a transcription factor that is required for the interaction of ESR1 with chromatin (Lupien et al., 2008; Hurtado et al., 2011). The deletion of *Foxa1* leads to early postnatal lethality (Bernardo et al., 2010). Therefore, to investigate the effects of *Foxa1* deletion in postnatal mammary gland development, the mammary rudiments (epithelium and surrounding stroma) from postnatal day 1 *Foxa1*-/- and control mice were grafted onto the kidney capsule of wild type recipient females (Bernardo et al., 2010). This grafting technique is used in mammary gland biology to study factors whose deletion causes early (embryonic or postnatal) death. These experiments demonstrated that FOXA1 is required for TEB formation and ductal elongation, because contrary to the grafted wild type mammary epithelium, the *Foxa1*-/- mammary epithelia fail to form TEBs and the rudimentary ducts are unable to extend into the mammary fat pad in response to host-derived pubertal signals (Bernardo et al., 2010). In the wild type mammary gland, *Foxa1* is expressed in $CD61^{pos}$ luminal progenitors cells, and is further upregulated in mature luminal cells, lending support to the hypothesis that FOXA1 contributes to the specification of the luminal lineage (Bernardo et al., 2010). The established role of *Prlr*, *Pgr*, *Esr1*, *Areg* and *Foxa1* in ductal morphogenesis, and their enriched expression in $Sca1^{pos}$ cells supports the hypothesis that $Sca1^{pos}$ cells contain ductal lineage-primed progenitors.

Further support for ductal priming in Sca1^{pos} cells is provided by the functional annotation clustering analysis of the microarray data (Fig. 2.8). All of the differentially expressed genes that passed the FDR cutoff of ≤0.01 (776 genes) were used to generate functional clusters using the DAVID Bioinformatics Resource. The differentially expressed genes in Sca1^{pos} cells are generally enriched in functions related to hormone response, ductal growth, branching and tube formation (Appendix 2). For example, cluster 1 contains *Cd24*, which is a known marker of mammary luminal epithelial cells (Sleeman et al., 2006), and high expression of CD24 is associated with high ESR1 expression (Sleeman et al., 2007). CD24 also regulates mammary ductal branching, as in its absence, branching morphogenesis is accelerated during puberty and pregnancy (Cremers et al., 2010). Cluster 3 contains the gene encoding hypoxia inducible factor 1A (*Hif1a*). Although a role in mammary gland development has not been reported for *Hif1a*, a role in branching morphogenesis was recently demonstrated in the lung, where over-expression of a normoxia-stable isoform of *Hif1a* led to aberrant pulmonary branching, as well as to defects in the maturation of epithelial cells (Bridges et al., 2012). *Wnt4* is also included in cluster 3, and has a known role in progesterone-driven mammary ductal branching both during puberty and pregnancy (Brisken et al., 2000).

Interestingly, pleckstrin (cluster 2, e.g., *Rasgrf1*) and tetraspanin (cluster 6, e.g., *Tspan13*) homology were also highly enriched in the ductal gene set. The pleckstrin homology (PH) domain is found in a wide range of proteins involved in intracellular signalling and cytoskeletal function (Haslam et al., 1993), such as the Rho GTPase RASGRF1 (Buchsbaum et al., 1996). The PH domain allows proteins to associate with specific membranes within the cell by binding to phosphatidylinositol lipids, thereby positioning them to interact with components of their respective signalling pathways. RASGRF1 (also known as P190) is associated with the plasma membrane, and functions as a downstream effector of IGF signalling (Heckman et al., 2007). Deletion of *P190b* leads to defects in mammary bud formation during embryonic development (Heckman et al., 2007). *P190b*-/- mammary epithelia also fail to express steroid receptors and display defects in TEB formation (Heckman et al., 2007), whereas over-expression of *P190b* leads to increased branching, delayed

ductal elongation and perturbed organization of the mammary tree (Vargo-Gogola et al., 2006). The tetraspanins are four transmembrane domain proteins that are associated with laminin-binding integrins (e.g., α6β1), and are known to regulate cell proliferation, cell motility and tumour invasiveness (Novitskaya et al., 2010). Recently, the tetraspanin family member TSPAN13 was implicated in TGF-β-mediated mammary tumour progression, with a possible role in tumour-stromal interactions during mammary tumour cell migration and invasion (Matise et al., 2012). As well, the up-regulation of CD151, another tetraspanin family member, was associated with advanced stage mammary ductal carcinoma *in situ* (34 % of advanced tumours showed up-regulation of CD151) (Novitskaya et al., 2010). Deletion of *Cd151* from an *in vitro* model of ductal carcinoma *in situ* (HB2 cell line) led to the formation of an internal lumen and polarization of the epithelial cells in the colonies (Novitskaya et al., 2010).

The exon arrays also show that 160 genes are more highly expressed in Sca1^{neg} cells compared to Sca1^{pos} cells. Included in this gene list are the previously reported genes encoding lactoferrin [LTF, (Sleeman et al., 2007)], which is an important iron-binding milk protein, as well as other milk components, such as *Csn3* and *Xdh*. *Xdh* is expressed uniformly in most cell types, and encodes a rate-limiting enzyme in purine metabolism (Vorbach et al., 2002). An additional function for XDH was demonstrated in the mammary gland, where it is required for the secretion of milk fat droplets from alveolar cells during lactation (Vorbach et al., 2002). Some milk genes, shown by others to be more highly expressed in Sca1^{neg} luminal cells (*Csn2*, *Mfge8*, *Wap*), did not pass the false detection rate filter of ≤0.01 in our experiments. Although not significant, qRT-PCR showed that *Csn2* is more highly expressed in Sca1^{neg} cells (Fig. 2.6), which is consistent with Sleeman's observations (Sleeman et al., 2007). The gene encoding the transcription factor ELF5, an important regulator of alveolar development, also displayed higher expression in Sca1^{neg} cells. ELF5 is a key mediator of prolactin signalling during pregnancy and lactation, and a block in alveolar differentiation is observed in *Elf5*-/- mammary glands, whereas precocious alveolar differentiation and milk secretion is observed when *Elf5* is over-expressed (Oakes et al., 2008). Overall, enriched expression of the alveolar genes *Ltf*, *Csn3*,

Xdh and *Elf5* in Sca1^{neg} cells supports the hypothesis that this population contains alveolar-primed luminal progenitors.

Functional annotation clustering analyses provide further support for the alveolar lineage priming in Sca1^{neg} cells. During pregnancy, the mammary gland undergoes significant remodeling to form secretory alveolar cells in preparation for lactation. Oakes described this process as the 'alveolar switch', whereby prolactin and progesterone signalling initiate a coordinated change in the genetic programming of the mammary epithelium to induce cell proliferation, migration and differentiation (Oakes et al., 2006). Consistent with these functions, Sca1^{neg} cells are enriched in genes associated with secretion, cell migration, extracellular matrix interactions and ion transport (Fig. 2.8B, Appendix 3). For example, cluster 1 and cluster 3 contain the gene encoding connective tissue growth factor (CTGF), a matrix-associated protein that is required for the lactogenic differentiation of cultured mammary epithelial cells (Morrison et al., 2010). Over-expression of *Ctgf* led to increased differentiation of mammary epithelial cells *in vitro* (using the HC11 mammary epithelial cell line) as demonstrated by increased *Csn2*, AKT and CCND1 expression (Wang et al., 2008; Morrison et al., 2010). On the other hand, depletion of CTGF by small interfering RNA (siRNA) knockdown led to decreased differentiation of the mammary epithelial cells (Wang et al., 2008). The authors posit that the changes in the expression of extracellular matrix proteins, integrins and *Csn2* due to CTGF over-expression/depletion suggest that CTGF is an important mediator of the stromal-epithelial communication that regulates alveolar differentiation.

An interesting cluster that emerged from the functional annotation analysis contains genes associated with the plexin-semaphorin signalling axis (cluster 7 in Fig. 2.8B). Semaphorins are secreted or membrane-associated glycoproteins that bind their cognate receptors, the plexins (Janssen et al., 2010). The plexin-semaphorin signalling axis is involved in neuronal axon guidance, vascular growth and tumour progression, primarily through the regulation of Rho-family GTPases (Kruger et al., 2005). *Met*, a gene we found to be enriched in Sca1^{neg} cells, and a member of cluster 7 in Figure 2.8B, was recently implicated in the regulation of mammary gland

growth, repopulating potential, and basal cell-fate commitment (Gastaldi et al., 2012). Consistent with our findings, Gastaldi also found that *Met* is more highly expressed in ESR1^{neg} luminal progenitors, and further showed that constitutive activation of MET leads to de-differentiation of luminal progenitor cells whereby they attain enhanced *in vitro* clonogenic and *in vivo* repopulating potential (Gastaldi et al., 2012). These data demonstrate a mechanism for the observation that basal-like breast cancer originates from luminal progenitor cells (Gastaldi et al., 2012).

Interestingly, the expression of the number of genes involved in the alveolar program is not as numerous in Sca1^{neg} cells as those in the ductal program in Sca1^{pos} cells, as evidenced by the lower level of enrichment of alveolar genes, as well as the lower number of genes in the A-gene signature (Fig. 2.7). One explanation for this is that the hormone-responsive ductal epithelium develops early in the lifecycle of the mammary gland, and therefore a higher basal level of ductal gene expression is observed. On the other hand, the alveolar cells mature later in development (pregnancy and lactation), and therefore the basal level of alveolar gene expression may be lower. This is supported by the observation that *Csn2* is expressed at much lower levels in luminal cells compared to *Nkcc1* and *Esr1* (Fig. 2.6).

In summary, the mouse exon arrays and qRT-PCR analyses demonstrate that the Sca1^{pos} and Sca1^{neg} luminal compartments contain cells primed towards ductal and alveolar cell fates, respectively. The Sca1^{pos} (ductal) gene signature is enriched in hormone signalling, ductal morphogenesis and tube formation, whereas the Sca1^{neg} (alveolar) gene signature is enriched in secretion, cell motility and extracellular matrix interactions.

2.5.3 In vitro functional output reflects in vivo lineage priming of luminal progenitor cells

To determine whether the lineage priming we observed *in vivo* affects the *in vitro* differentiation potential of Sca1^{pos} and Sca1^{neg} cells, they were cultured under differentiation conditions. The first culturing conditions included prolactin and

RANKL for the last 7 days to promote the differentiation of alveolar and ductal progenitor cells (Fig. 2.9). PRL signalling is essential for the proliferation, terminal differentiation and survival of alveolar cells (Brisken et al., 1999; Gallego et al., 2001), whereas RANKL promotes ductal side-branching and alveolar differentiation *in vivo* (Fernandez-Valdivia et al., 2009). Both progenitor populations form smooth, round colonies in the absence of differentiation factors (Fig. 2.9A, left panels). After culturing the colonies with PRL/RANKL for 7 days, however, $Sca1^{pos}$ colonies change shape by forming protrusions that appear to be early stages of branching, whereas the morphology of the $Sca1^{neg}$ colonies does not change (Fig. 2.9A, right panels). These observations demonstrate that $Sca1^{pos}$ and $Sca1^{neg}$ cells respond differently to differentiation signals, and suggest that their lineage priming may impact their *in vitro* differentiation potential. Following the differentiation protocol, the D- and A-gene sets were analyzed by qRT-PCR to determine whether the *in vivo* lineage priming is maintained. The results presented in Figure 2.9B show that alveolar genes are more highly expressed in cultured $Sca1^{neg}$ cells compared to $Sca1^{pos}$ cells, and conversely, ductal genes are more highly expressed in cultured $Sca1^{pos}$ cells compared to $Sca1^{neg}$ cells, demonstrating that the initial priming is preserved.

To further show that *in vitro* branching morphogenesis is restricted to $Sca1^{pos}$ cells, we cultured $Sca1^{pos}$ and $Sca1^{neg}$ cells using differentiation factors (EGF/FGF/heparin) that promote ductal branching (Hung et al., 2005; Pavlovich et al., 2011). Consistent with the results we obtained previously by culturing the cells with PRL and RANKL, only $Sca1^{pos}$ cells formed branched colonies when cultured with EGF/FGF/heparin (Fig. 2.10). The $Sca1^{pos}$ colonies generated using this differentiation protocol have elongated, tubular, highly branched projections. Some colonies were also completely tubular (top right), whereas others had branched projections originating from a round colony (top left). These morphologies may reflect progressive stages of differentiation. $Sca1^{pos}$ cells also form more colonies (~22 %) than $Sca1^{neg}$ cells (~10 %). This result is consistent with Li's published observation that approximately 20 % of $Sca1^{pos}$ luminal progenitor cells and 12 % of $Sca1^{neg}$ luminal progenitor cells have colony forming activity (Li et al., 2009).

Approximately 15 % of Sca1^{pos} cells form branched colonies, compared to 1.6 % of Sca1^{neg} cells, demonstrating that branching morphogenesis is limited to Sca1^{pos} cells. However, 85 % of colonies formed by Sca1^{pos} cells do not generate branches, and remain round and smooth despite the addition of tubulogenic factors. Heterogeneity within the ductal luminal compartment has been documented based on ESR1 expression (LaMarca and Rosen, 2008). However, to my knowledge, ductal luminal progenitors have not been functionally dissected before. These data demonstrate that Sca1^{pos} cells contain (at least) two different progenitor populations: one that engages in branching morphogenesis, and one that cannot. These observations are of interest, and future investigations will focus on the molecular differences between branching and non-branching Sca1^{pos} cells.

In summary, this chapter described the isolation of two luminal progenitor cell populations (Sca1^{pos} and Sca1^{neg} cells) and showed that they are differentially primed in their gene expression towards ductal and alveolar cell fates, respectively. Furthermore, we showed that the *in vivo* genetic priming affects the *in vitro* differentiation potential of Sca1^{pos} and Sca1^{neg} luminal progenitor cells. Together, these findings led to the development of a revised model of the cellular hierarchy that maintains the adult mammary gland (Fig. 2.11). As shown by others, the Ma-CFC contains a bipotent luminal progenitor/stem cell population, which thus far has been described as CD29^{lo}, CD24^{hi} and CD49f^{lo} (Stingl et al., 2006; Shackleton et al., 2006), and additionally as K8+/K19+ (Van Keymeulen et al., 2011). This progenitor/stem cell, through one or more undefined intermediates, gives rise to two luminal cell lineages: ductal and alveolar (Fig. 2.11A). The characterization of two new progenitor cell populations, as described in this chapter, provides a critical link between the bipotent luminal progenitor/stem cells and the two differentiated luminal cells types in the adult mammary gland (Fig. 2.11B).

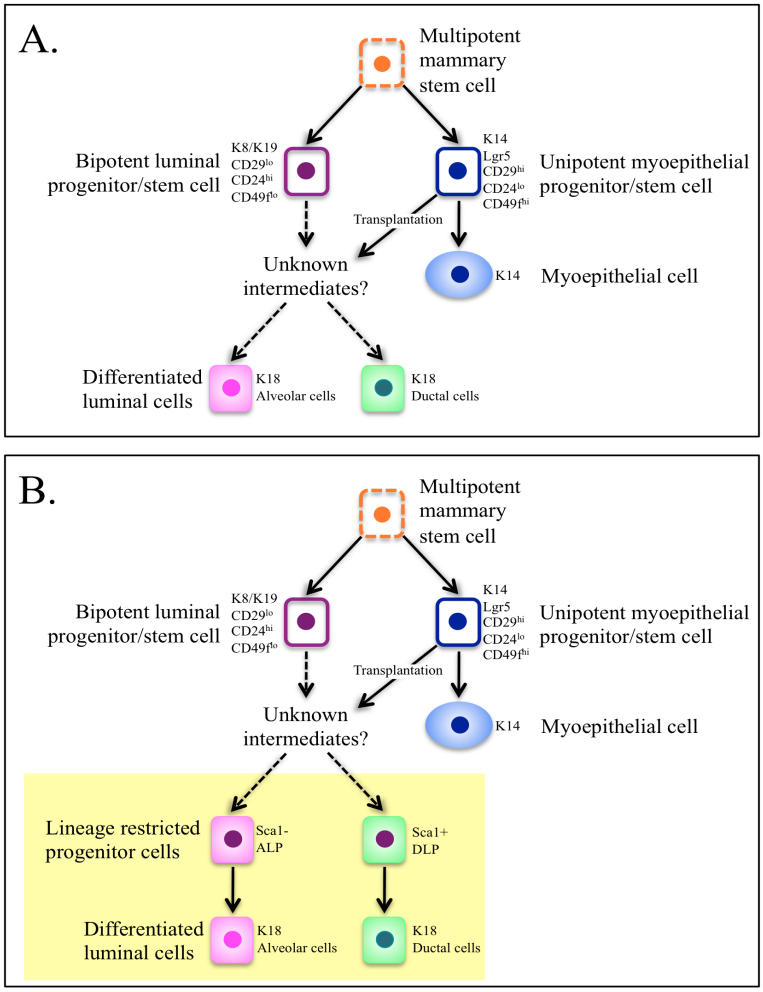


Figure 2.11 Schematic representation of the cellular hierarchy in the mammary gland. A. In this model, the adult mammary gland is maintained by a multipotent stem cell population (shown in orange), which, based on Stingl and Shackleton's experiments, resides in the myoepithelial cell compartment. This multipotent stem cell gives rise to two lineage-restricted stem/progenitor cells (bipotent luminal progenitor/stem cells [referred to as Ma-CFC in Chapter 1] and unipotent myoepithelial progenitor/stem cells [referred to as MRU in Chapter 1]). These, in turn, give rise to the three principal mammary epithelial cell types: myoepithelial cells, ductal cells and alveolar cells. In a transplantation setting, the myoepithelial progenitor/stem cell generates both primary cell lineages (myoepithelial and luminal), whereas the bipotent luminal progenitor/stem cell only generates luminal cell types. As indicated by discontinuous arrows, the intermediate cell types between the bipotent luminal progenitor/stem cell and differentiated luminal cells were not known. **B.** In this model, the two progenitor cell types identified in this chapter (Sca1- ALPs and Sca1+ DLPs) provide a link between the bipotent luminal progenitor/stem cell and the differentiated luminal cell types (in yellow box). Key: continuous arrows- relationship described; discontinuous arrow- relationship unknown; discontinuous line around cell-identity contended.

3 C/EBPβ is required for alveolar luminal progenitor programming

3.1 Introduction

The development of the mammary gland requires the specification of myoepithelial, ductal and alveolar lineages from mammary stem and progenitor cells. Others have shown that the luminal epithelial compartment contains progenitor/stem cells (Shackleton et al., 2006; Stingl et al., 2006; Van Keymeulen et al., 2011) that can be divided into two populations: one that expresses hormone-sensing genes, and one that expresses alveolar genes (Sleeman et al., 2007). In Chapter 2, we have further shown that improved separation of luminal progenitor cells using a novel combination of cell surface markers can isolate two lineage-primed progenitor populations, where $Sca1^{pos}$ luminal cells contain ductal luminal progenitors (DLPs), and $Sca1^{neg}$ luminal cells contain alveolar luminal progenitors (ALPs). However, the genetic mechanisms that regulate the specification of these luminal cell fates and their subsequent differentiation remains poorly understood.

Three CCAAT/enhancer binding protein family members are expressed in the mammary gland: C/EBPα, -β, and -δ. C/EBPα is not required for mammary gland development, as demonstrated by transplantation experiments (Seagroves et al., 1998), whereas C/EBPδ is important for the regulation of apoptotic genes during involution (Thangaraju et al., 2005). These studies suggested that neither of these two C/EBP family members are involved in the early specification of the two luminal cell types. Expression of *Cebpb*, on the other hand, is low in the virgin gland, increases during pregnancy, declines during lactation, and increases again 24-48 hours after the onset of involution (Gigliotti and DeWille, 1998; Robinson et al., 1998; Sabatakos et al., 1998). C/EBPβ has also been shown to be an important regulator of mammary gland development (Robinson et al., 1998; Seagroves et al., 1998). In adult *Cebpb*-/- mice, mammary ducts are morphologically abnormal

(distended) with fewer side-branches, and alveolar differentiation is blocked in response to lactogenic hormones in transplanted *Cebpb*-/- mammary epithelia (Robinson et al., 1998; Seagroves et al., 1998). Further studies have shown that these defects are intrinsic to the mammary epithelium, as a lack of C/EBPβ in the stroma does not affect ductal elongation/branching during puberty, or development/ differentiation of the alveoli during pregnancy (Grimm and Rosen, 2003). Based on these observations, we posited that C/EBPβ may be involved in the specification of luminal cells, and subsequently, in regulating the development of alveolar cells.

3.2 Aims of this chapter

This chapter aims to investigate the role of C/EBPβ in the transcriptional diversification of ductal and alveolar cell fates in the luminal compartment of the mammary gland. To begin, flow cytometric analysis was performed on C/EBPβ-deficient and control mammary glands to assess the composition of the luminal compartment. This was followed by gene profiling using qRT-PCR on C/EBPβ-deficient $Sca1^{pos}$ and $Sca1^{neg}$ luminal cells to assess any potential changes in the D- and A- gene signatures. Lastly, C/EBPβ-deficient $Sca1^{pos}$ and $Sca1^{neg}$ luminal cells were cultured *in vitro* to investigate any changes in proliferative activity and differentiation potential (preliminary data only).

3.3 Collaborators

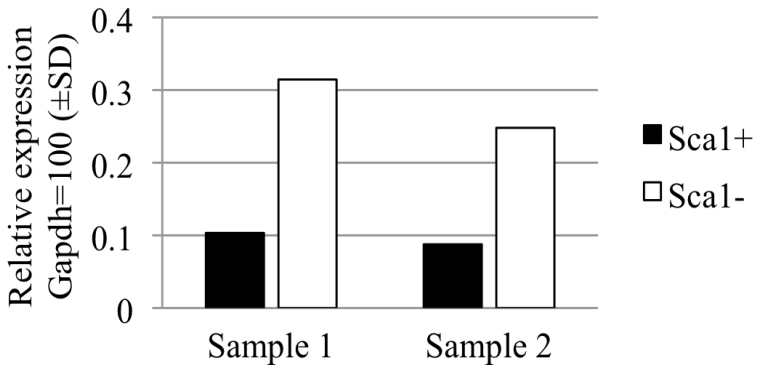
The work described in this chapter was performed in collaboration with by Dr. Susana Garcia-Silva. Dr. Silva performed the qRT-PCR from which Figure 3.8 was generated. I performed the other experiments presented in this chapter, and generated all of the figures.

3.4 Results

3.4.1 C/EBPβ affects the cellular organization of the mammary gland

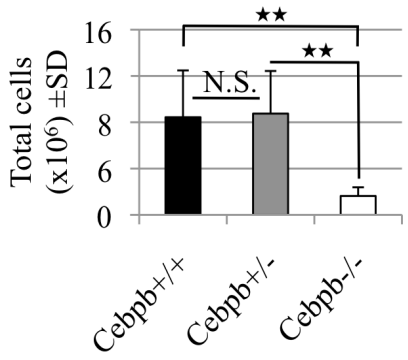
The level of *Cebpb* expression was previously shown to be low in the virgin mammary gland (Sabatakos et al., 1998). However, its expression in specific mammary gland cell populations has not been assessed. Therefore, the level of *Cebpb* expression was first determined in wild type $Sca1^{pos}$ and $Sca1^{neg}$ cells by qRT-PCR. The level of *Cebpb* expression in adult (10-12 week old) wild type females is approximately 3-fold higher in $Sca1^{neg}$ cells compared to $Sca1^{pos}$ cells (Fig. 3.1). This observation suggests that C/EBPβ may be important in priming $Sca1^{neg}$ cells towards alveolar development. To test this, I used the flow cytometric staining strategy described in Figure 2.5 to assess potential changes in the cellular composition of the luminal compartment in *Cebpb*-/- mice. The total cellularity of *Cebpb*-/- mammary glands (10-12 weeks old) is approximately 5-fold lower than littermate controls ($p<0.005$) (Fig. 3.2), consistent with the proliferative defects in *Cebpb*-/- mammary glands observed by others *in vivo* (Robinson et al., 1998; Seagroves et al., 1998). In the luminal compartment ($Lin^{neg}CD29^{pos}$ $CD24^{hi}CD49f^{lo}$), a shift from $Sca1^{neg}$ to $Sca1^{pos}$ cells is observed in *Cebpb*-/- mice when compared to *Cebpb*+/- or *Cebpb*+/+ littermates (Fig. 3.3). As well, the previously published observations that alveolar proliferation and differentiation are significantly affected in the absence of C/EBPβ are consistent with these findings.

Absolute quantification of the luminal (Ma-CFC, $Sca1^{pos}$, $Sca1^{neg}$), myoepithelial (MRU&MYO) and stromal cells shows that all three populations are significantly decreased in *Cebpb*-/- females (Fig. 3.4 & 3.5). Relative quantification, defined here as frequency of Lin^{neg}, also shows that luminal cells are increased by approximately 20 %, with a concomitant decrease in myoepithelial (~35 %) and stromal cells (~45 %) when compared to wild type littermates (Fig. 3.6). The luminal and myoepithelial cells are similar in frequency between *Cebpb*+/+ and *Cebpb*+/- mice. However, the stromal compartment of *Cebpb*+/- is reduced compared to *Cebpb*+/+ to a frequency more similar to *Cebpb*-/-, suggesting that C/EBPβ may be important for the



Cebpb ±SD	Sca1+ (n=2)	Sca1- (n=2)
Relative expression Gapdh=100	0.096 ±0.011	0.281 ±0.047

Figure 3.1 Relative expression of *Cebpb* in ductal (Sca1^{pos}) and alveolar (Sca1^{neg}) luminal progenitor cells. Quantitative reverse transcriptase PCR analysis of *Cebpb* expression was performed on sorted ductal and alveolar luminal progenitor cells from 12-week old wild type virgin females. The values represent the average *Cebpb* expression normalized to *Gapdh* from two biological replicates (Sample 1 and Sample 2), each assayed in technical triplicate. **Abbreviations:** SD, standard deviation; n, number.



Total cells (x10^6) ±SD	Cebpb+/+ (n=3)	Cebpb+/- (n=4)	Cebpb-/- (n=5)
	8.44±4.04	8.75±3.68	1.63±0.75
	0.4594		0.0018
	N.S.		
	0.0042		

Figure 3.2 Total cellularity in the mammary gland of *Cebpb*-/- mice compared to littermate controls. Bar graph and accompanying table showing the total number of mammary epithelial cells in *Cebpb*+/+ (n=3), *Cebpb*+/-, (n=4) and *Cebpb*-/- (n=5) mice. The values in the bar graph represent the mean number of cells (x10^6) calculated from a minimum of three biological replicates (precise number is indicated for each genotype in the table). The values in the bar graph represent the mean, and error bars indicate standard deviation. Significance (p-value) was calculated using an unpaired, one-tailed Student's T-test. The level of significance is shown by stars in the graphs, and numerically in the tables. **Abbreviations:** SD, standard deviation; N.S., not significant.

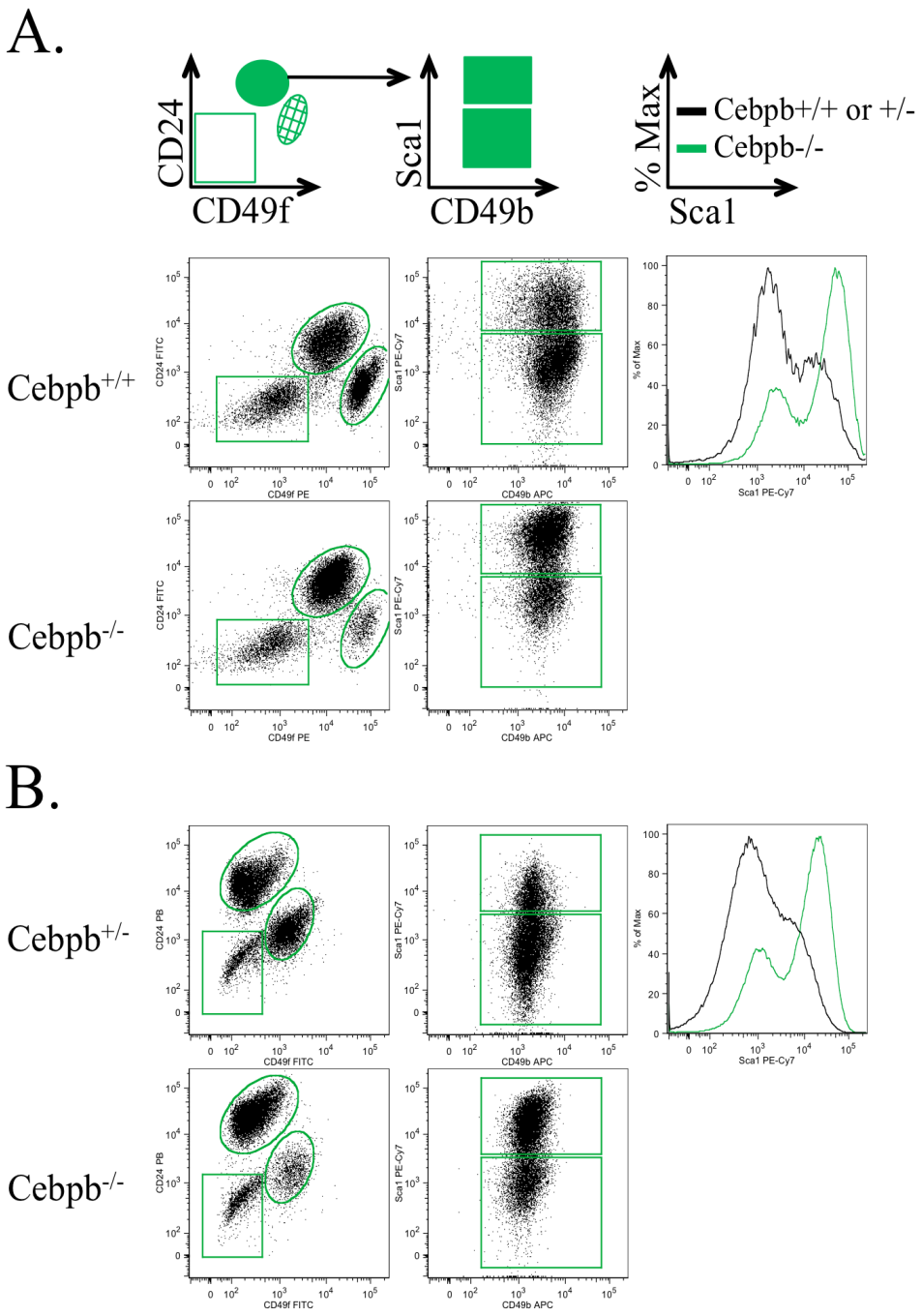


Figure 3.3 Deletion of *Cebpb* alters the cellular composition of the mammary luminal compartment. (A) A schematic diagram is shown in the top row and illustrates the gating strategy used in the representative FACS dot plots below. The CD24/CD49f plot was generated by first gating on singlets, live cells, lineage (CD31, CD45, Ter119) negative cells, and CD29 positive cells (not shown). In the histograms, black lines represent control samples, and green lines represent *Cebpb*-/- samples. Comparison between *Cebpb*+/+ and *Cebpb*-/- is shown in (A), and comparison between *Cebpb*+/- and *Cebpb*-/- is shown in **(B)**.

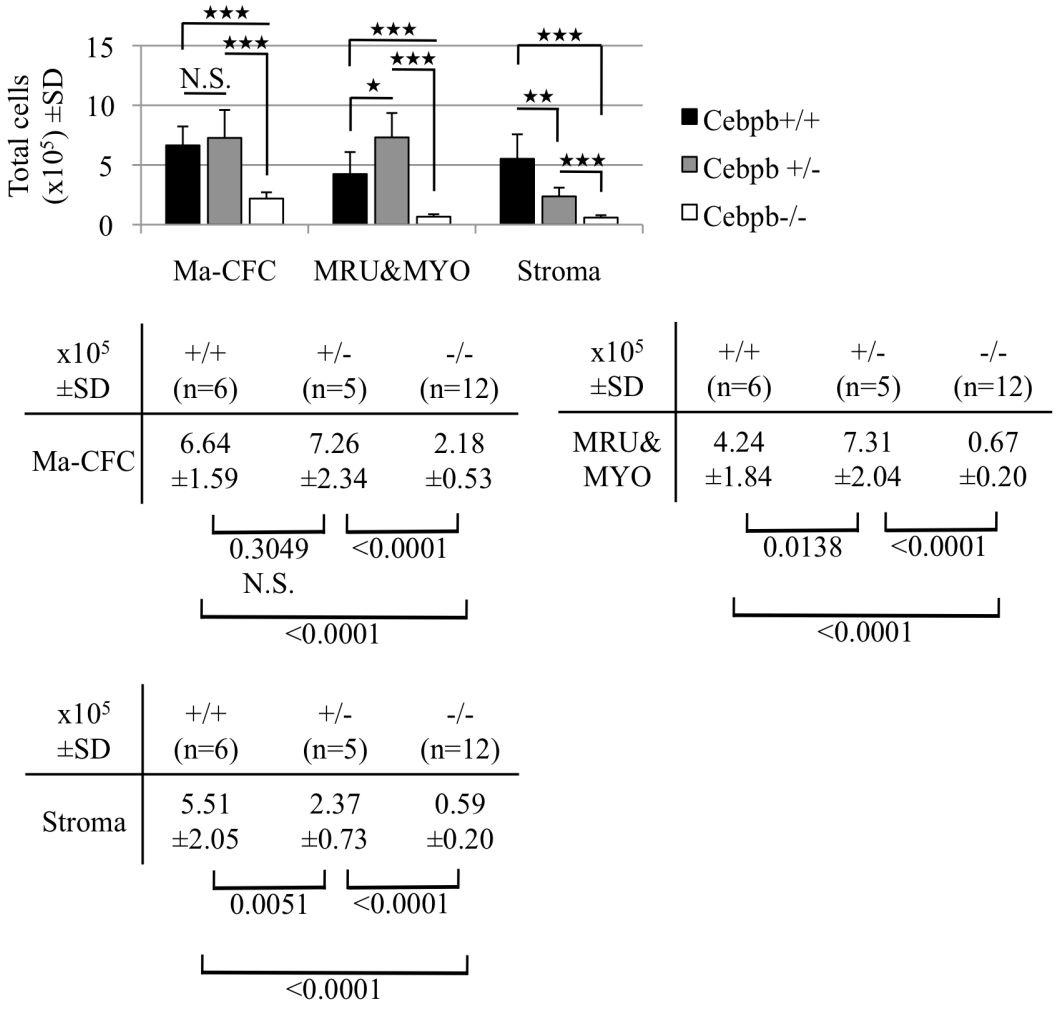
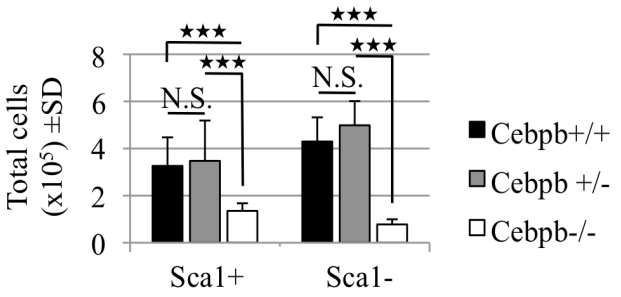


Figure 3.4 Absolute quantification of the luminal, myoepithelial and stromal cell compartments of the mammary gland in *Cebpb-/-* mice and littermate controls. The three cell populations in the CD24/CD49f plots generated as in Fig.3.3A were quantified by multiplying the frequency of total cells for each population by the average total number of cells for each genotype. The bar graph and accompanying tables show the mean number of cells for each population. The number of biological replicates used to generate each mean is shown in the tables. Error bars indicate standard deviation. Significance (p-value) was calculated using an unpaired, one-tailed Student's T-test. The level of significance is shown by stars in the graphs, and numerically in the tables below. **Abbreviations:** SD, standard deviation; N.S., not significant; Ma-CFC, luminal cells; MRU&MYO, basal cells (mammary repopulating units and myoepithelial cells).



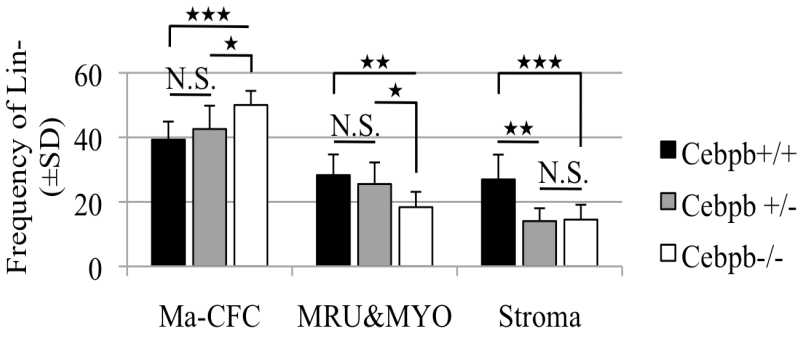
x10^5 ±SD	+/+ (n=6)	+/- (n=5)	-/- (n=12)
Sca1+	3.26 ±1.21	3.48 ±1.71	1.35 ±0.32

0.4065 | 0.0003
N.S.
<0.0001

x10^5 ±SD	+/+ (n=6)	+/- (n=5)	-/- (n=12)
Sca1-	4.30 ±1.03	4.99 ±1.03	0.78 ±0.22

0.1482 | <0.0001
N.S.
<0.0001

Figure 3.5 Absolute quantification of Sca1^{pos} and Sca1^{neg} luminal cells in *Cebpb-/-* mice and littermate controls. The two cell populations in the Sca1/CD49b plots generated as in Fig.3.3A were quantified by multiplying the frequency of total cells for each population by the average total number of cells for each genotype. The bar graph and accompanying tables show the mean number of cells for each population. The number of biological replicates used to generate each mean is shown in the tables. Error bars indicate standard deviation. Significance (p-value) was calculated using an unpaired, one-tailed Student's T-test. The level of significance is shown by stars in the graphs, and numerically in the tables below. **Abbreviations:** SD, standard deviation; N.S., not significant.



% Lin- ±SD	+/+ (n=6)	+/- (n=5)	-/- (n=10)
Ma-CFC	39.28 ±5.61	42.58 ±7.23	50.03 ±4.37

0.2079 0.0130
N.S.
0.0004

% Lin- ±SD	+/+ (n=6)	+/- (n=5)	-/- (n=10)
MRU& MYO	28.30 ±6.40	25.54 ±6.68	18.36 ±4.75

0.2514 0.0155
N.S.
0.0016

% Lin- ±SD	+/+ (n=6)	+/- (n=5)	-/- (n=10)
Stroma	26.97 ±7.69	14.04 ±3.98	14.50 ±4.62

0.0041 0.4257
N.S.
0.0006

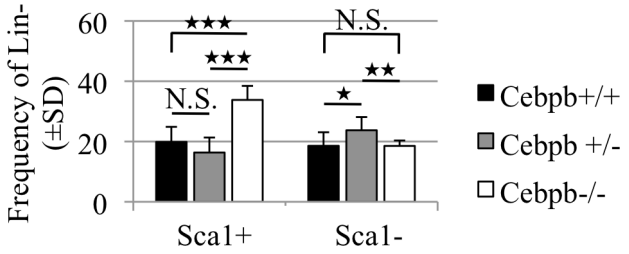
Figure 3.6 Relative quantification of the luminal, myoepithelial and stromal cell compartments of the mammary gland in *Cebpb-/-* mice and littermate controls. The three cell populations in the CD24/CD49f plots generated as in Fig.3.3A were quantified as frequency of Lin (CD31, CD45, Ter119) negative cells. The bar graph and accompanying tables show the mean frequency of cells for each population. The number of biological replicates used to generate each mean is shown in the tables. Error bars indicate standard deviation. Significance (p-value) was calculated using an unpaired, one-tailed Student's T-test. The level of significance is shown by stars in the graphs, and numerically in the tables below. **Abbreviations:** SD, standard deviation; N.S., not significant; Ma-CFC, luminal cells; MRU&MYO, basal cells (mammary repopulating units and myoepithelial cells).

development of the stroma. Within the luminal compartment, the frequency of Sca1^{pos} cells is significantly increased in *Cebpb-/-*, whereas no change is observed in the frequency of Sca1^{neg} cells, except that heterozygous mice appear to have a higher frequency of Sca1^{neg} cells when compared to *Cebpb*+/+ (p=0.04) and *Cebpb-/-* (p=0.003) mice (Fig. 3.7). These analyses demonstrate that C/EBPβ is required for appropriate specification of the ductal and alveolar lineages, and in its absence, the Sca1^{pos}/Sca1^{neg} ratio is increased significantly, suggesting that C/EBPβ may specifically promote alveolar development.

3.4.2 C/EBPβ controls luminal progenitor specification

To determine if the perturbed ductal and alveolar specification resulting from genetic deletion of *Cebpb* is caused by changes in the gene expression of Sca1^{pos} and Sca1^{neg} cells, qRT-PCR was performed using the D- and A- gene lists that were developed in Chapter 2 (Fig. 2.7). Gene profiling of *Cebpb*-/- Sca1^{neg} cells shows a general increase in D-genes with concomitant decrease in A-genes (Fig. 3.8), suggesting a loss of alveolar lineage priming in these cells. In Sca1^{pos} cells, the residual alveolar programming is depleted, whereas ductal genes further enriched, albeit at low levels. Common luminal genes (*Krt18*, *Sox9*, *Klf5* and *Gata3*) are slightly down-regulated in both populations.

The change in the Sca1^{pos}/Sca1^{neg} ratio in *Cebpb-/-* mice, and the increase in ductal priming in *Cebpb-/-* Sca1^{neg} cells suggest that the functional program of these cells may be affected. To investigate this, *Cebpb-/-* Sca1^{pos} and Sca1^{neg} cells were cultured in semisolid medium under conditions that promote ductal morphogenesis (heparin, FGF, EGF) to assess if Sca1^{neg} cells have the ability to form branched structures *in vitro*. Preliminary data (n=1 for mutant and control) presented in Figure 3.9 shows that *Cebpb-/-* Sca1^{neg} cells form irregularly shaped colonies (Fig. 3.9B) compared to the smooth, round colonies formed by the wild type control (Fig. 3.9A). However, *Cebpb-/-* Sca1^{neg} cells did not form branches like wild type ductal progenitor cells. As well, *Cebpb-/-* Sca1^{pos} cells did form some irregularly shaped colonies (Fig. 3.9B), but displayed a reduced level of branching compared to controls.



% Lin- ±SD	+/+ (n=6)	+/- (n=5)	-/- (n=10)
Sca1+	19.87 ±5.00	16.35 ±4.98	33.80 ±4.66

0.1370 <0.0001
N.S.
<0.0001

% Lin- ±SD	+/+ (n=6)	+/- (n=5)	-/- (n=10)
Sca1-	18.57 ±4.50	23.73 ±4.38	18.54 ±1.8

0.0438 0.0027
0.4912
N.S.

Figure 3.7 Relative quantification of Sca1^{pos} and Sca1^{neg} luminal cells in *Cebpb-/-* mice and littermate controls. The two cell populations in the Sca1/CD49b plots generated as in Fig.3.3A were quantified as frequency of Lin (CD31, CD45, Ter119) negative cells. The bar graph and accompanying tables show the mean frequency of cells for each population. The number of biological replicates used to generate each mean is shown in the tables. Error bars indicate standard deviation. Significance (p-value) was calculated using an unpaired, one-tailed Student's T-test. The level of significance is shown by stars in the graphs, and numerically in the tables below. Supplementary information related to this figure ('frequency of parent' analysis of Sca1^{pos} and Sca1^{neg} luminal cells) is shown in Appendix 4. **Abbreviations:** SD, standard deviation; N.S., not significant.

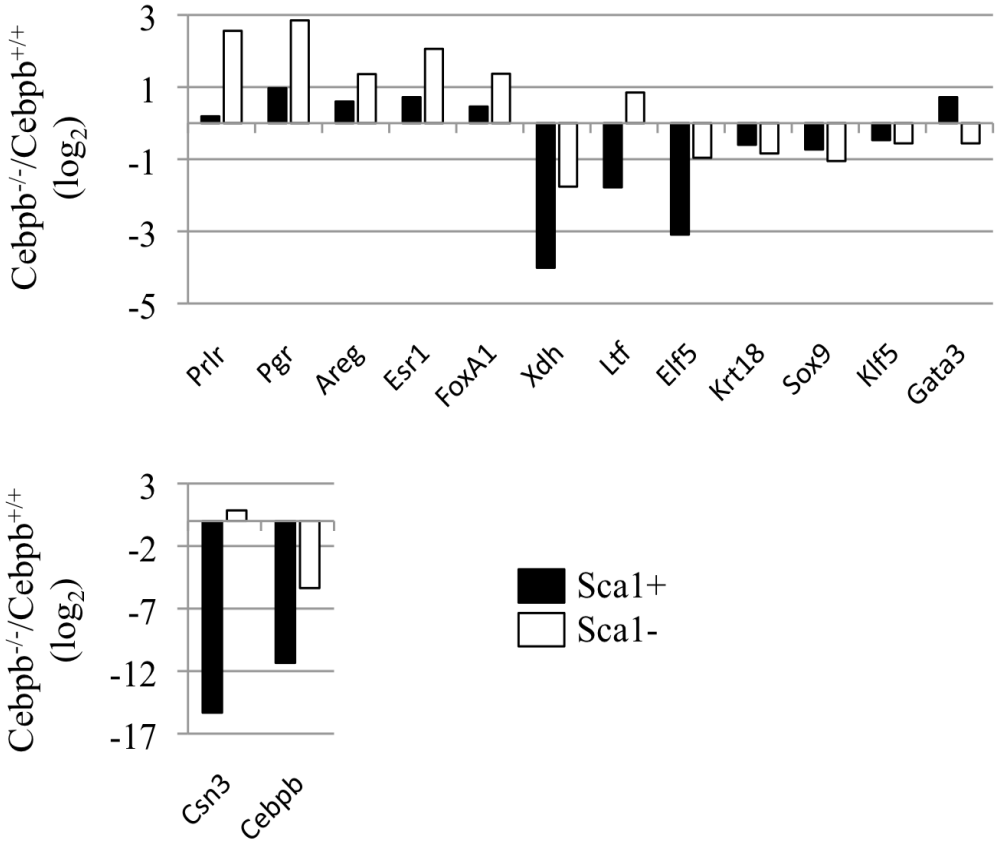


Figure 3.8 The effect of *Cebpb* deletion on gene expression in Sca1^{pos} and Sca1^{neg} cells. Gene expression analysis comparing ductal (*Prlr*, *Pgr, Areg*, *Esr1*, *Foxa1*), alveolar (*Xdh*, *Ltf*, *Elf5*, *Csn3*) and common luminal genes (*Krt18*, *Klf5*, *Gata3*) from sorted *Cebpb*-/- and *Cebpb*+/+ Sca1^{pos} and Sca1^{neg} luminal cells. Gene expression was determined from three biological replicates for each genotype, and analyzed in triplicate by qRT-PCR using Fluidigm 48.48 dynamic arrays. The values shown are the ratio of the average Sca1^{pos} and Sca1^{neg} values after $\log_2$ transformation.

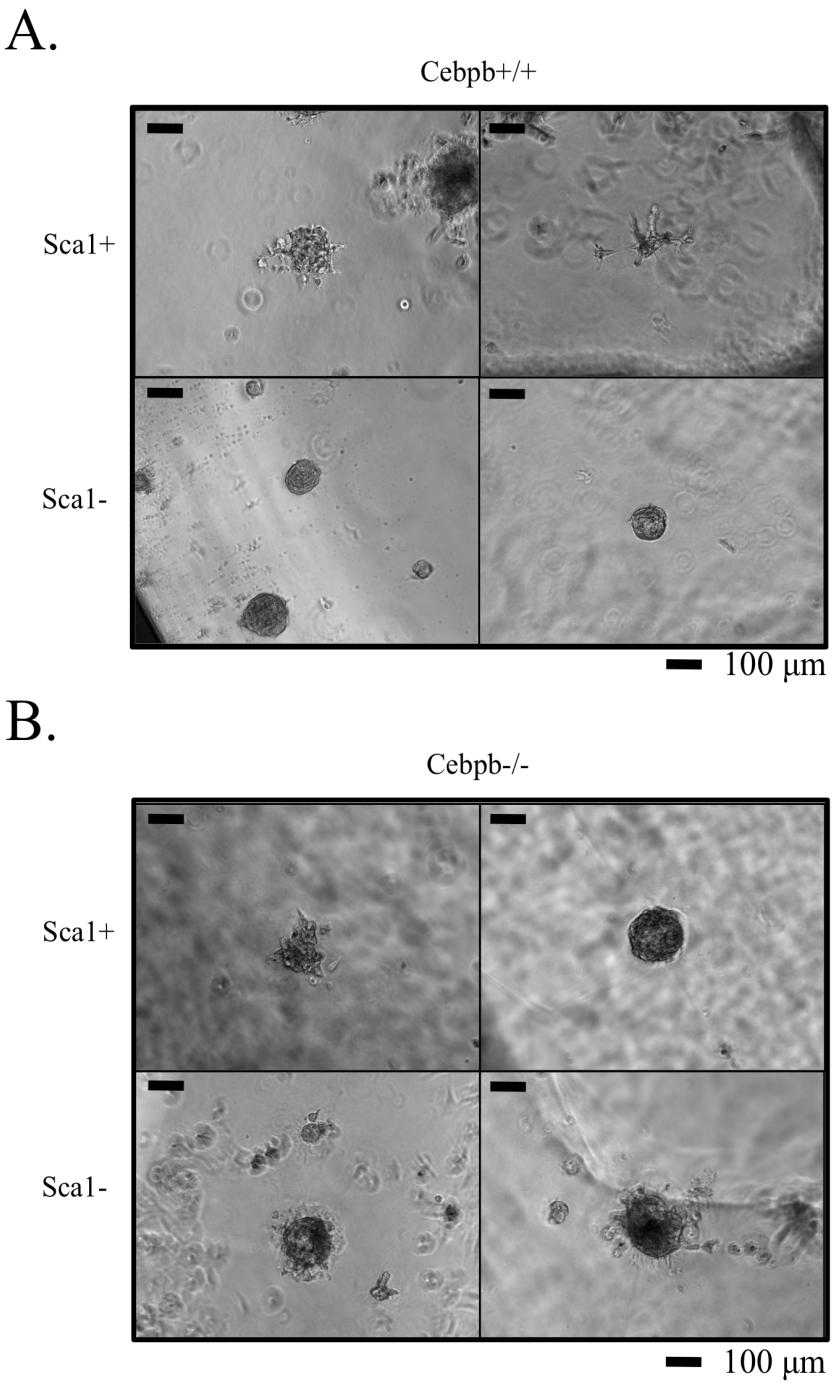


Figure 3.9 The effect of *Cebpb* deletion on *in vitro* colony morphology. Representative images of $Sca1^{pos}$ and $Sca1^{neg}$ colonies embedded in Matrigel, and cultured over a layer of mitomycin-C treated NIH3T3 fibroblast cells in the presence of EGF, FGF and heparin for 21 days. Scale bar=100 µm. **(A)** Colonies generated from sorted *Cebpb*+/+ (n=1) $Sca1^{pos}$ and $Sca1^{neg}$ cells. $Sca1^{pos}$ colonies are irregularly shaped with some colonies forming branches, whereas $Sca1^{neg}$ colonies are smooth and round. **(B)** Colonies generated from *Cebpb*-/- (n=1) $Sca1^{pos}$ and $Sca1^{neg}$ cells. Some $Sca1^{pos}$ colonies are irregularly shaped but lack distinct branches, while most are round. Most $Sca1^{neg}$ colonies are round, but not smooth, as in (A).

These data suggest that deletion of *Cebpb* affects the functional output of Sca1^{pos} and Sca1^{neg} cells, but additional experiments are required to replicate these data, and to assess the molecular changes that cause these defects.

3.5 Discussion

3.5.1 C/EBPβ is important for alveolar lineage priming

The previous chapter described the isolation of two luminal progenitor cell populations (Sca1^{pos} and Sca1^{neg} cells) and showed that they are differentially primed in their gene expression towards ductal and alveolar cell fates, respectively. As well, *in vivo* lineage priming was found to affect the *in vitro* differentiation potential of Sca1^{pos} and Sca1^{neg} cells. However, the genetic mechanisms that regulate the specification of these luminal cell fates and their subsequent differentiation is not well understood. Recently, several studies began to investigate the role of transcription factors (e.g., GATA3, ELF5) in mammary gland development. GATA3 first emerged as an important marker of breast cancers of luminal origins (Charafe-Jauffret et al., 2005). This finding prompted further investigation into the possible roles of GATA3 during normal mammary gland development, and it was subsequently shown that GATA3 is an essential regulator of luminal cell differentiation (Asselin-Labat et al., 2006). Targeted (K14) embryonic deletion of *Gata3* blocks the formation of mammary placodes and nipple epithelia. As well, in the adult, loss of GATA3 in the luminal compartment leads to reduced ductal elongation and branching, as well as reduced alveolar development during pregnancy. This decline in differentiated cell types is coupled with an increase in CD61^{pos} luminal progenitors, demonstrating that GATA3 is required for luminal cell differentiation. ELF5 was also shown to be important for alveolar differentiation (Oakes et al., 2008), and more recently, to play a critical role in specifying the luminal cell fate and regulating progenitor/stem cell function (Chakrabarti et al., 2012). Like *Gata3*, conditional deletion of *Elf5* from the mammary gland leads to a block in luminal progenitor cell specification, as shown by the loss of CD61^{pos} luminal cells and concomitant increase in K14/K8 double positive cells. On the other

hand, over-expression of *Elf5* causes precocious alveolar development, including milk secretion, in virgin females, demonstrating that ELF5 specifies alveolar cell fate in luminal progenitor cells. Based on these findings, and the published observations that deletion of *Cebpb* also leads to defects in alveologenesis and ductal development, we decided to investigate the role of C/EBPβ in luminal lineage specification more closely. Although the expression of *Cebpb* is low in the virgin mammary gland, its expression is higher in alveolar progenitors than in ductal progenitors (Fig. 3.1). This observation suggests that C/EBPβ may be involved in the priming of alveolar luminal progenitor cells, similar to ELF5, which could be important for their subsequent lineage commitment. Similar observations in hematopoietic stem/progenitor cells, where genetic priming defines lineage potential in later developmental stages, has been described (Mancini et al., 2011). In Mancini's study, for example, FOG1, a cofactor of GATA proteins, was shown to prime hematopoietic stem cells (HSCs) towards the megakaryocytic/ erythroid (Mk/E) cell fate, and in its absence, HSCs could not generate Mk/E lineage committed progenitors. This study demonstrates that early genetic priming by transcription/co-factors, which can precede observable changes in gene expression in precursor cells, is required for subsequent lineage commitment.

To investigate whether C/EBPβ is involved in lineage priming of luminal progenitors, we assessed the composition of the luminal compartment utilizing our new staining strategy, which allowed us to specifically assess C/EBPβ function in ductal and alveolar-primed progenitor cells. As previously described, C/EBPβ is essential for the proliferation and survival of mammary epithelial cells (Robinson et al., 1998; Seagroves et al., 1998; 2000), and consistent with these previous studies, we also found that the total cellularity of *Cebpb*-/- mammary glands is greatly reduced (Fig. 3.2). The dot plots also revealed a prominent change in the luminal compartment of *Cebpb*-/- mice (Fig. 3.3). Consistent with the known role of C/EBPβ in alveolar development, we observed that alveolar-phenotypic cells are greatly reduced in *Cebpb*-/- mice. Also, the reduction of alveolar-phenotypic cells is coupled with an increase in ductal-phenotypic cells, suggesting that normal alveolar differentiation may occur through specification of ductal cells ($Sca1^{pos}$) into alveolar

cells ($Sca1^{neg}$). The relative increase in the frequency of total luminal progenitor cells (Fig. 3.6) suggests that C/EBPβ is important for maintaining the balance between luminal and myoepithelial cell fates, as well as for the correct specification of each luminal lineage, as deletion of *Cebpb* leads to alterations in the frequency of $Sca1^{pos}$ and $Sca1^{neg}$ luminal cells (Fig. 3.7, Appendix 4). The analysis in Figure 3.7 shows that $Sca1^{pos}$ cells increase in *Cebpb*-/- mice compared to controls, and Appendix 4 shows more clearly that the relative frequency of $Sca1^{neg}$ cells decreases in *Cebpb*-/- compared to controls. The latter observation is not clear in the analysis in Fig. 3.7, because the increase in the frequency of total luminal cells masks the relative decrease in $Sca1^{neg}$ luminal cells. The analysis in Figure 3.7 was performed in this way, rather than as frequency of the parent population, to ensure consistency across all analyses included in this thesis, including those in Chapter 2, and to allow for direct comparison across all data sets. In this instance, these data reveal that in the whole mammary epithelium, *Cebpb*-/- $Sca1^{pos}$ cells are more numerous than in controls, whereas $Sca1^{neg}$ cells are of similar frequency. However, when the higher frequency of total luminal cells is taken into account (Appendix 4), it becomes more evident that there is a relative reduction in $Sca1^{neg}$ cells and relative increase in $Sca1^{pos}$ cells, which increases the $Sca1^{pos}/Sca1^{neg}$ ratio from approximately 1.1:1 in control mice to approximately 2.5:1 in *Cebpb*-/- mice. When these experiments were performed, the effect of *Cebpb* deletion in luminal progenitor cells was not known. However, a study by LaMarca published in 2010 included very similar observations to these (LaMarca et al., 2010). Her investigations also demonstrated that deletion of *Cebpb* leads to accumulation of $CD24^{lo}Sca1^{pos}$ and reduction of $CD24^{lo}Sca1^{neg}$ luminal cells (LaMarca et al., 2010). As well, the block in the differentiation of luminal progenitor cells that we observed (Fig. 3.6), as shown by the relative increase in Ma-CFCs (luminal cells) and the relative decrease in MRU&MYO (myoepithelial/basal cells), was also shown in LaMarca's study.

To determine if the changes in the organization of the luminal compartment in *Cebpb*-/- mice are accompanied by changes in the lineage gene signatures of phenotypic ductal and alveolar cells, the expression of the representative D- and A-lineage genes was assessed by qRT-PCR. Figure 3.8 shows that all D-lineage genes

tested were up-regulated, and all but one (*Ltf*) A-lineage gene tested were down-regulated in both *Cebpb*-/- Sca1^{pos} and Sca1^{neg} cells. As well, expression of the common luminal genes were also down-regulated. Further to these results, LaMarca showed that the expression of differentiation markers (e.g., *Eya1*, *Notch3*) were also down-regulated in *Cebpb*-/- luminal cells, whereas stem cell genes were up-regulated (e.g., *Pbx1*, *K15*, *K5*) (LaMarca et al., 2010). When taken together with the observation that the relative frequency of phenotypic ductal cells is higher in *Cebpb* -/- mice than controls, these data demonstrate that alveolar lineage programming is dependent on the presence of C/EBPβ.

The data presented in Figs. 2.7, 2.9 and 2.10 demonstrated that lineage priming affects *in vitro* differentiation potential. Therefore, to assess the functional impact of the changes in the expression of A-lineage and D-lineage genes that were observed in *Cebpb*-/- mice, *Cebpb*-/- Sca1^{pos} and Sca1^{neg} luminal cells were cultured in semisolid medium. The experiment presented in Figure 3.9 is only a preliminary finding. However, initial observations suggest that the colony morphology of *Cebpb* -/- Sca1^{pos} and Sca1^{neg} luminal cells is altered when compared to controls. As well, although the colony counts are not presented here, the general impression is that the number of colonies is greatly reduced in *Cebpb*-/- cells when compared to *Cebpb*+/+ cells. This observation is consistent with LaMarca's findings, who noted that while approximately 7 % of luminal cells from *Cebpb*+/+ mice give rise to colonies, only 2 % of luminal cells from *Cebpb*-/- are able to do so (LaMarca et al., 2010). LaMarca also noted that *Cebpb*-/- cells generate smaller colonies than *Cebpb*+/+ cells. In the experiment shown in Figure 3.9, an overt difference in colony size is not observed. However, additional experiments may reveal similar observations to LaMarca's. Follow-up experiments to this study will also assess the gene expression pattern of cultured Sca1^{pos} and Sca1^{neg} cells to determine if the changes in lineage priming are maintained in culture, as was seen for wild type cells (Fig. 2.7).

In summary, the changes in the cellular organization of the mammary gland upon deletion of *Cebpb* demonstrate that C/EBPβ is required for appropriate specification of the two luminal cell fates. As well, C/EBPβ promotes alveolar lineage priming, as

in its absence, alveolar genes are down-regulated and ductal genes are up-regulated in both Sca1^{pos} and Sca1^{neg} luminal cells. Lastly, a preliminary *in vitro* colony forming assay shows that in addition to severe proliferation defects, the changes in *in vivo* lineage priming in *Cebpb*-/- mice may also affect the *in vitro* differentiation potential of *Cebpb*-/- Sca1^{pos} and Sca1^{neg} cells.

4 Identification of C/EBPβ target genes in the mouse mammary gland

4.1 Introduction

Cebpb-/- mice exhibit mammary gland developmental defects in ductal morphogenesis, and alveolar proliferation and differentiation. Overt defects are not observed in prepubertal females, however by 8-12 weeks of age, the ducts are morphologically abnormal with fewer branches (Robinson et al., 1998; Seagroves et al., 1998). Alveolar development is also disturbed, as transplanted *Cebpb*-/- mammary epithelia form significantly fewer alveoli than controls in response to lactogenic hormones. However, the alveoli that form appear histologically normal (Robinson et al., 1998; Seagroves et al., 1998). Comparison of gene expression arrays of luminal cells ($Lin^{neg}CD24^{hi}CD29^{lo}$) in *Cebpb*-/- and control mice demonstrate that Notch and Wnt/β-catenin signalling pathways are significantly altered in *Cebpb*-/- cells (LaMarca et al., 2010). As well, the expression of basal cell markers, such as *K5*, *K14*, *p63*, are also increased in *Cebpb*-/- luminal cells, suggesting that specification of the luminal cell fate may be perturbed in the absence of C/EBPβ (LaMarca et al., 2010). Data presented in Chapter 3 illustrates that C/EBPβ is specifically required for alveolar lineage commitment, and its absence is marked by disturbed organization of the luminal compartment, as well as up-regulation of ductal genes and concomitant down-regulation of alveolar genes in both alveolar and ductal luminal progenitor cells. Investigations into the mechanisms of C/EBPβ-mediated lineage specification and commitment have been hampered by the lack of known target genes in the mammary gland. To date, no direct *in vivo* targets have been identified for C/EBPβ in the mammary gland, and *in vitro* studies have identified only four mammary genes (*Id2*, *Csn2*, *Xdh* and *Prlr*) as direct C/EBPβ targets (Raught et al., 1995; Kim et al., 2002; Karaya et al., 2005; Seymour et al., 2006; Goldhar et al., 2011).

4.2 Aims of this chapter

The aim of this chapter is to identify direct C/EBPβ target genes in the mammary gland *in vivo*. The data presented in Chapter 3, together with LaMarca's observations (LaMarca et al., 2010), convincingly demonstrate that C/EBPβ is required for luminal lineage commitment and subsequent programming towards alveolar cell fate. However, the molecular mechanisms underlying C/EBPβ action remain to be elucidated, and the lack of known C/EBPβ target genes in the mammary gland has hampered further investigations. Therefore, to identify the direct target genes of C/EBPβ *in vivo*, a chromatin immunoprecipitation sequencing (ChIP-Seq) protocol was developed and optimized using freshly isolated mammary luminal cells from virgin mice ($Sca1^{pos}$ and $Sca1^{neg}$ cells) and mice at day 16.5 of pregnancy (P16.5).

4.3 Collaborators

Dr. Simon Tomlinson curated the microarray and ChIP-Seq data, and generated the lists of differentially expressed genes. Dr. Susana Garcia-Silva prepared the virgin luminal cells and I prepared the pregnant luminal cells for microarray. Susan Moore performed the luciferase assay presented in Figure 4.14. I performed all other experiments and analyses presented in this chapter. The bioinformatics tool used to generate some of the figures (GeneProf) was developed by Florian Halbritter (Halbritter et al., 2012).

4.4 Results

4.4.1 Gene expression comparison between differentiated alveolar cells and two luminal progenitors

My interest in investigating the role of C/EBPβ in alveolar programming prompted me to include a second time point in alveolar development (day 16.5 of pregnancy, denoted as P16.5). Choosing this second developmental time point late in late pregnancy enabled the investigation of C/EBPβ function in differentiated alveolar

cells that have not yet entered the secretory phase. The flow cytometric strategy used to isolate mature alveolar cells at P16.5 ($Lin^{neg}CD29^{pos}CD24^{hi}CD49f^{lo}Sca1^{neg}$) for microarray analysis and ChIP-Seq is shown in Figure 4.1A. In P16.5 mammary glands, the $Sca1^{pos}$ luminal cells (last plot to the right) almost completely disappear. The loss of $Sca1^{pos}$ luminal cells and concomitant increase in $Sca1^{neg}$ luminal cells is consistent with a shift from predominantly ductal to predominantly alveolar cells during late pregnancy (Brisken and O'Malley, 2010).

The differentially expressed gene set for P16.5 was generated by selecting the common differentially expressed genes from the comparisons between P16.5 and virgin $Sca1^{neg}$ cells, and P16.5 and virgin $Sca1^{pos}$ cells (FDR $\leq$0.01). P16.5 alveolar cells differentially express 1198 genes. When the changes in gene expression of the selected ductal and alveolar genes (from Chapter 2) are analyzed, they show that ductal genes are strongly down-regulated in P16.5 when compared to $Sca1^{pos}$ cells, and alveolar genes are strongly upregulated, except for *Ltf* (Fig. 4.1B). When compared to $Sca1^{neg}$ cells, the directions of change in gene expression are variable for both ductal and alveolar genes. For example, while *Pgr* and *Areg* are down-regulated in P16.5 compared to $Sca1^{neg}$ cells, *Prlr* and *Esr1* are more highly expressed (Fig. 4.1C). As well, the expression of the alveolar genes *Csn3*, *Xdh* and *Elf5* are further enriched, as expected, but *Ltf* is down-regulated. The common luminal genes in both comparisons are down-regulated. These data show a clear distinction in the gene expression of $Sca1^{pos}$ cells and P16.5 alveolar cells, demonstrating their unique lineage identity, whereas no clear trend is apparent between $Sca1^{neg}$ cells and P16.5, suggesting that they are more closely related cell types.

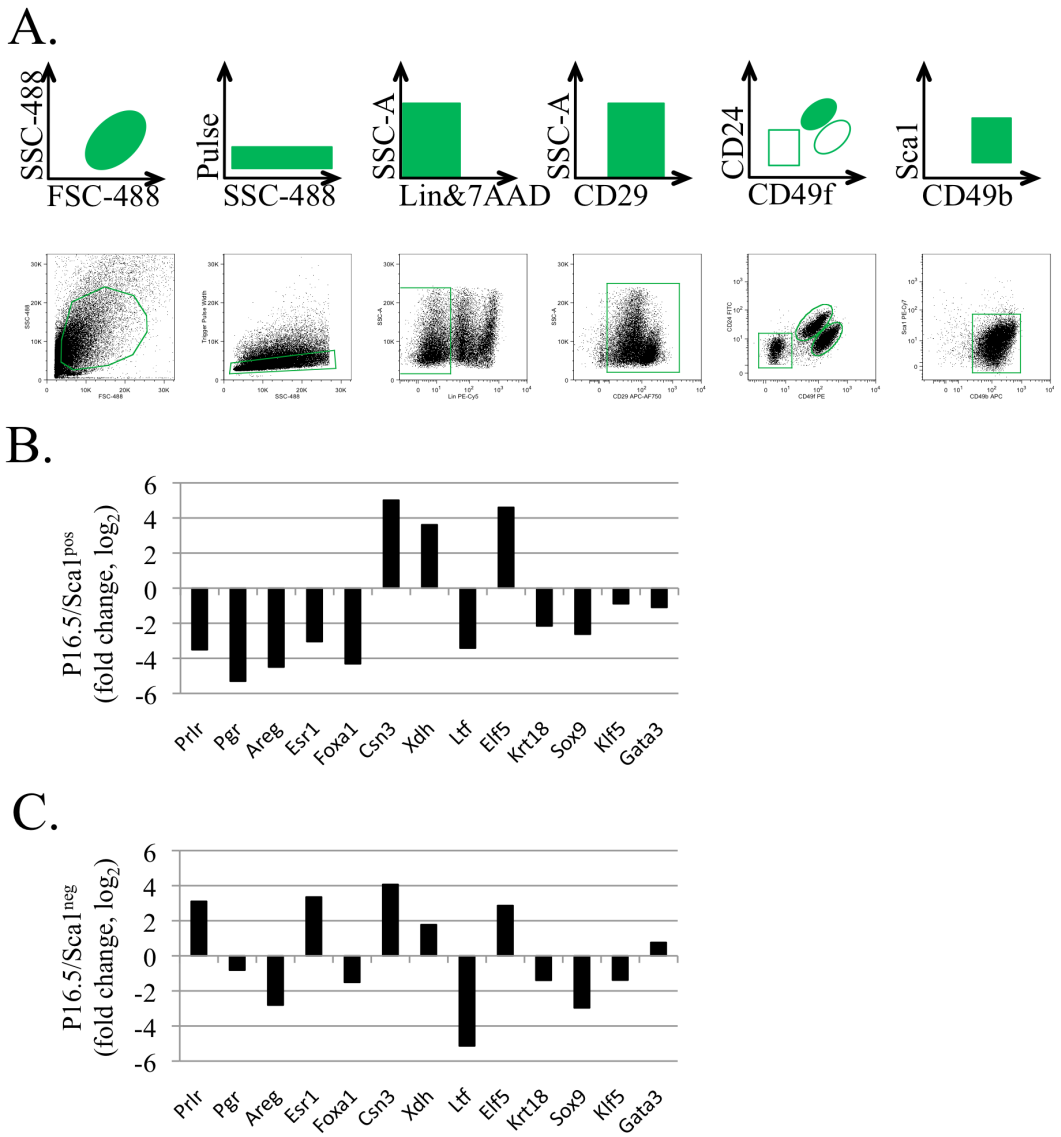


Figure 4.1 Flow cytometric profile and gene expression analysis of P16.5 (day 16.5 of pregnancy) luminal cells. (A) Schematic diagram (top row) and representative dot plots (bottom row) depicting the gating strategy used to isolate Sca1^{neg} luminal cells from P16.5 pregnant mice (termed as P16.5). Cells were first gated on scatter to obtain singlets. Then lineage (CD31, CD45, Ter119) positive cells were excluded, and Lin^{neg} cells were selected based on CD29 expression. CD29^{pos} cells were then evaluated for CD24 and CD49f expression. CD24^{hi}CD49f^{lo} cells were further subdivided based on Sca1 expression. **(B)** The expression of the ductal-specific and alveolar-specific genes identified in Figure 2.7 was assessed in P16.5 luminal cells. The normalized, annotated, filtered expression values are shown as the ratio between the average P16.5 (n=4) and virgin Sca1^{pos} (n=3) values after $\log_2$ transformation. **(C)** Same as in (B), except that P16.5 expression values were compared with virgin Sca1^{neg} rather than Sca1^{pos}. **Abbreviations:** SSC, side scatter; FSC, forward scatter; SSC-A, side scatter area; Lin, lineage; 7-AAD, 7-aminoactinomycin-D.

4.4.2 Optimization of C/EBPβ chromatin immunoprecipitation

To provide insight into how C/EBPβ regulates alveolar lineage specification, the genome-wide occupancy of C/EBPβ was investigated in ductal and alveolar progenitors, as well as in P16.5 alveolar cells using chromatin immunoprecipitation sequencing. An overview of the method is shown in Figure 4.2A. Following cell sorting as described previously (Chapter 2), the cells were fixed with 1 % formaldehyde, lysed and sonicated. Optimization of the SDS concentration of the sonication buffer is shown in (Appendix 5). To obtain fragmented chromatin with an average DNA length of 200-400 base pairs (bp), which is optimal for deep sequencing, different sonication conditions were tested (data not shown). The optimal sonication condition for mammary epithelial cells was determined to be 50 cycles of 30 seconds on/30 seconds off on a high setting using a Bioruptor Diagenode waterbath sonicator (Fig. 4.2B).

Next, in order to validate the chromatin immunoprecipitation experiment, a known C/EBPβ target gene was needed as a positive control. Due to the lack of known *in vivo* C/EBPβ target genes, *in silico* analysis (TRANSFAC) was used to identify putative C/EBPβ binding sites in the three published *in vitro* target genes (*Xdh*, *Id2*, *Csn2*) and an additional predicted target gene (*Gata3*) (Fig. 4.3A). These analyses were then used to design quantitative PCR (qPCR) assays (colored bars below gene schematics) against the genomic regions where C/EBPβ binding has been demonstrated (red triangles), or was expected based on the bioinformatics analyses (orange triangles). Quantitative PCR assays were also designed to screen unbound regions within the same gene to serve as intralocus negative controls. Quantitative PCR assays were successfully designed for screening bound regions in three of the four genes (*Xdh*, *Id2*, *Gata3*). However, although a qPCR assay was designed against the published binding sites in *Csn2*, this assay failed during initial testing, and because an alternate assay could not be designed in the same genomic region, it was excluded in subsequent experiments.

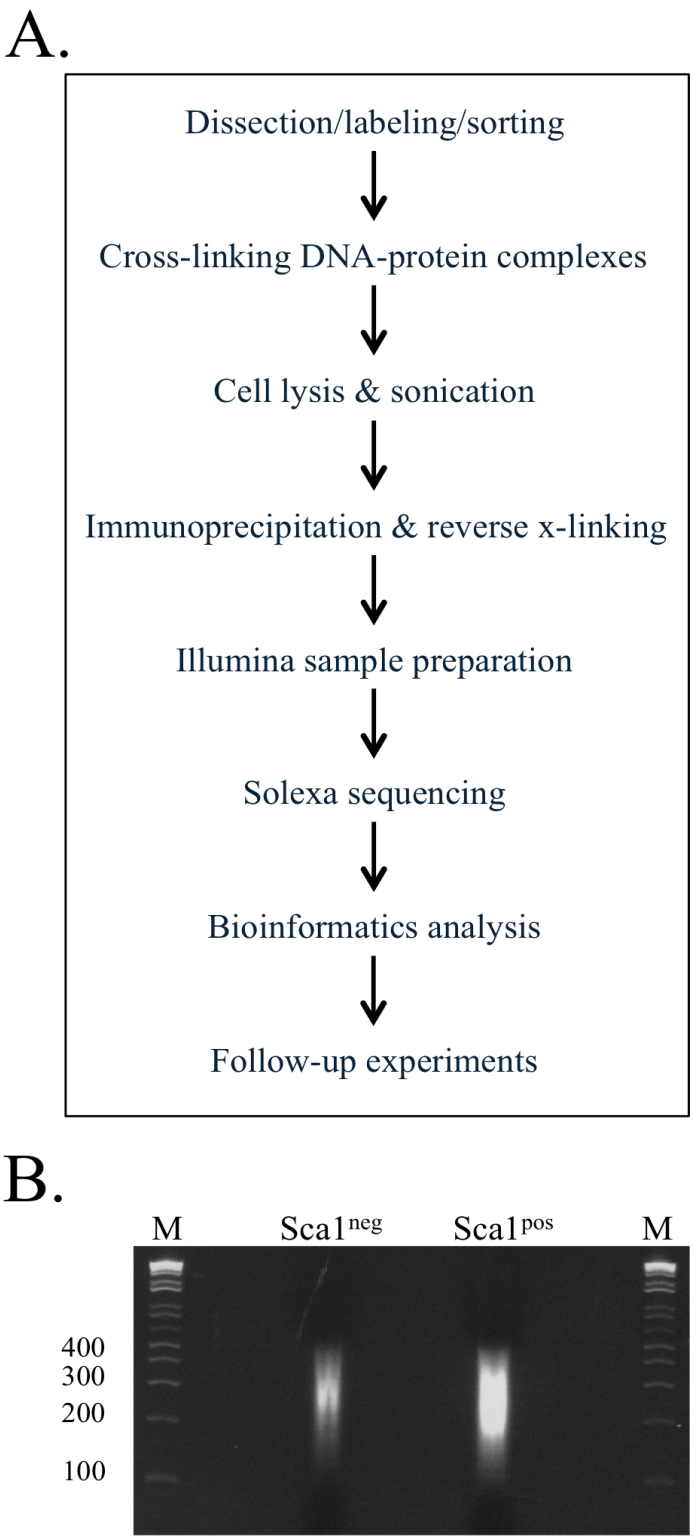


Figure 4.2 Overview of the C/EBPβ chromatin immunoprecipitation (ChIP) method. (A) Flow chart showing the steps in ChIP-Seq. **(B)** 2 % agarose gel shows the fragmented chromatin from sorted $Sca1^{neg}$ and $Sca1^{pos}$ luminal cells after optimization of sonication conditions. **Abbreviations:** M, marker.

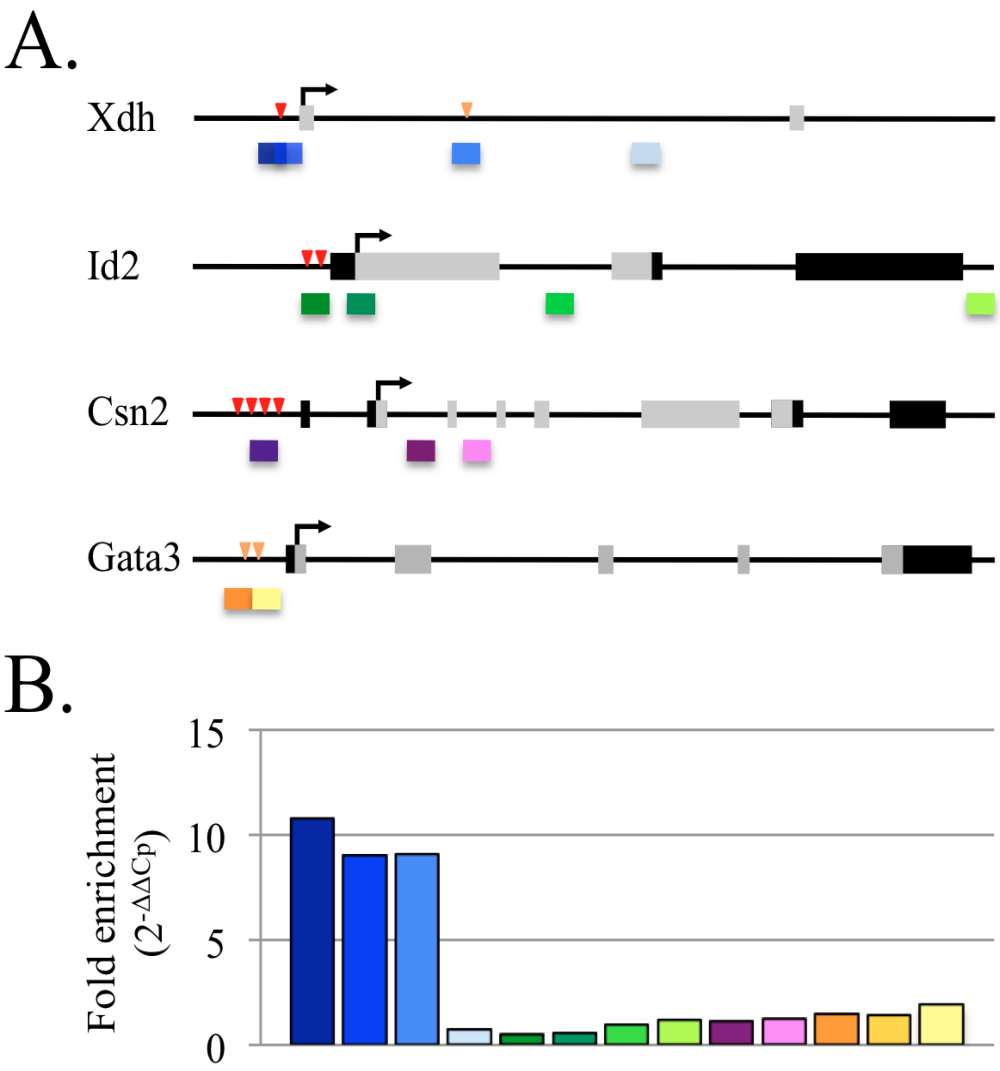
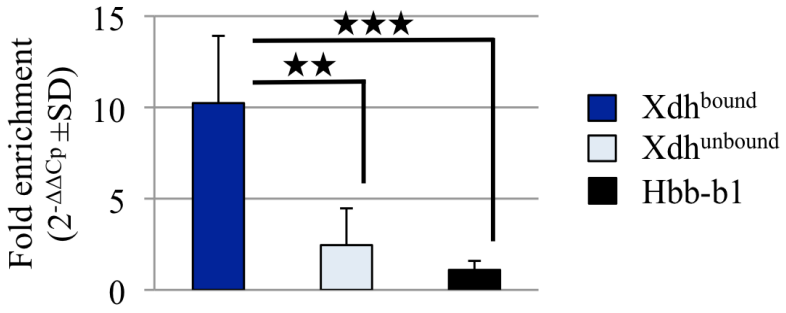


Figure 4.3 Screening of published C/EBPβ target genes in Sca1^{neg} luminal cells. (A) Schematic diagram (not to scale) showing the 3 published (*Xdh*, I*d2*, *Csn2*) and one putative (*Gata3*) target genes of C/EBPβ. The transcription start site is indicated by a black arrow for each gene. Exons are indicated as thicker lines, with black exon comprising untranslated regions and grey exons comprising the coding regions for each gene. Published (red arrow heads) and putative (orange arrow heads) C/EBPβ binding sites are indicated. Quantitative PCR assays for each gene are shown as colored bars below each gene schematic. **(B)** Quantitative PCR assessment of C/EBPβ ChIP using chromatin from sorted Sca1^{neg} luminal cells form 10-week old virgin females. Chromatin was immunoprecipitated with C19 anti-C/EBPβ or non-immune rabbit IgG. The bar graph represents fold enrichment between input-standardized C/EBPβ and IgG ChIP ($2^{-\Delta\Delta Cp}$). Colors correspond to the qPCR assays shown in (A).

The initial ChIP experiments were performed using Sca1^{neg} cells isolated from 10-week old virgin females. Each sample tube contained fragmented chromatin from 100,000 Sca1^{neg} luminal cells, and was immunoprecipitated using an anti-C/EBPβ antibody (C19) or control IgG (rabbit, of equal μg quantity to the anti-C/EBPβ antibody). The amount of antibody used in each sample tube was first titrated (Appendix 5) using the published binding sites in *Xdh*, and the optimal concentration (0.5 μg) was then used to assess binding in the other target genes (Fig. 4.3B). To calculate fold enrichment, the $2^{-\Delta Cp}$ and $2^{-\Delta\Delta Cp}$ methods were used, which compare the input-standardized Cp values of the C/EBPβ ChIP and the control IgG ChIP. Optimization of the washing and elution conditions was then conducted to increase the signal to noise ratio (data not shown). These experiments led to the observation that in Sca1^{neg} cells, C/EBPβ binds to the *Xdh* promoter, but not to the *Id2* or *Gata3* promoters (Fig. 4.3B). Binding in *Csn2* could not be determined because of the inability to design qPCR assays near the binding sites. The ChIP method was validated by 5 independent experiments, which reproducibly showed significant enrichment in the *Xdh* promoter (*Xdh*^{bound}) compared to the intralocus negative control (*Xdh*^{unbound}) and an interlocus negative control (*Hbb-b1* encoding β-globin) (Fig. 4.4).

An overview of the sequencing library sample preparation is shown in Figure 4.5A. Each sequencing library was prepared from 5 tubes of ChIP-ed DNA (2 ng of input material) and was verified by qPCR (Fig. 4.5C,D). The enrichment in C/EBPβ binding to the promoter of *Xdh* is highest in P16.5, followed by Sca1^{neg} cells then Sca1^{pos} cells, which correlates with increasing levels of *Xdh* expression in these cells (Fig. 4.5B). An example of an electropherogram from a sequencing library is shown in Figure 4.6A. Electropherograms were generated by the University of Edinburgh Genepool to verify the quality of each sequencing library. The quality of the sequencing libraries was assessed based on the distribution of DNA fragments, the size range and concentration of the DNA (Fig. 4.6B). All of the sequencing libraries displayed symmetrical distribution of the DNA fragments, had an appropriate size-range of 150-400 bp for Solexa sequencing, and had sufficient DNA concentration to pass the initial quality checks.

A.



B.

Fold enrichment (n=5)	Xdh bound	Xdh unbound	Hbb-b1
	10.23 ±3.68	2.45 ±2.01	1.09 ±0.49
	0.0016	0.0907	
		N.S.	
	0.0003		

Figure 4.4 C/EBPβ ChIP in Sca1^{neg} luminal cells from five independent experiments. (A) Quantitative PCR assessment of C/EBPβ ChIP using chromatin from sorted Sca1^{neg} luminal cells form 10-week old virgin females performed as in Fig. 4.3 from 5 independent experiments. The bar graph (E) represents fold enrichment between input-standardized C/EBPβ and IgG ChIP ($2^{-\Delta\Delta Cp}$). Error bars indicate standard deviations. P-values were calculated using an unpaired, one-tailed Student's T-test, and significance is shown as a star in the graph and as a value in the table below (F). Hbb-b1 (encoding β-globin) was used as an additional negative control. Supplementary information related to this figure (further optimization experiments) is shown in Appendix 5. **Abbreviations:** SD, standard deviation; n, number; N.S., not significant.

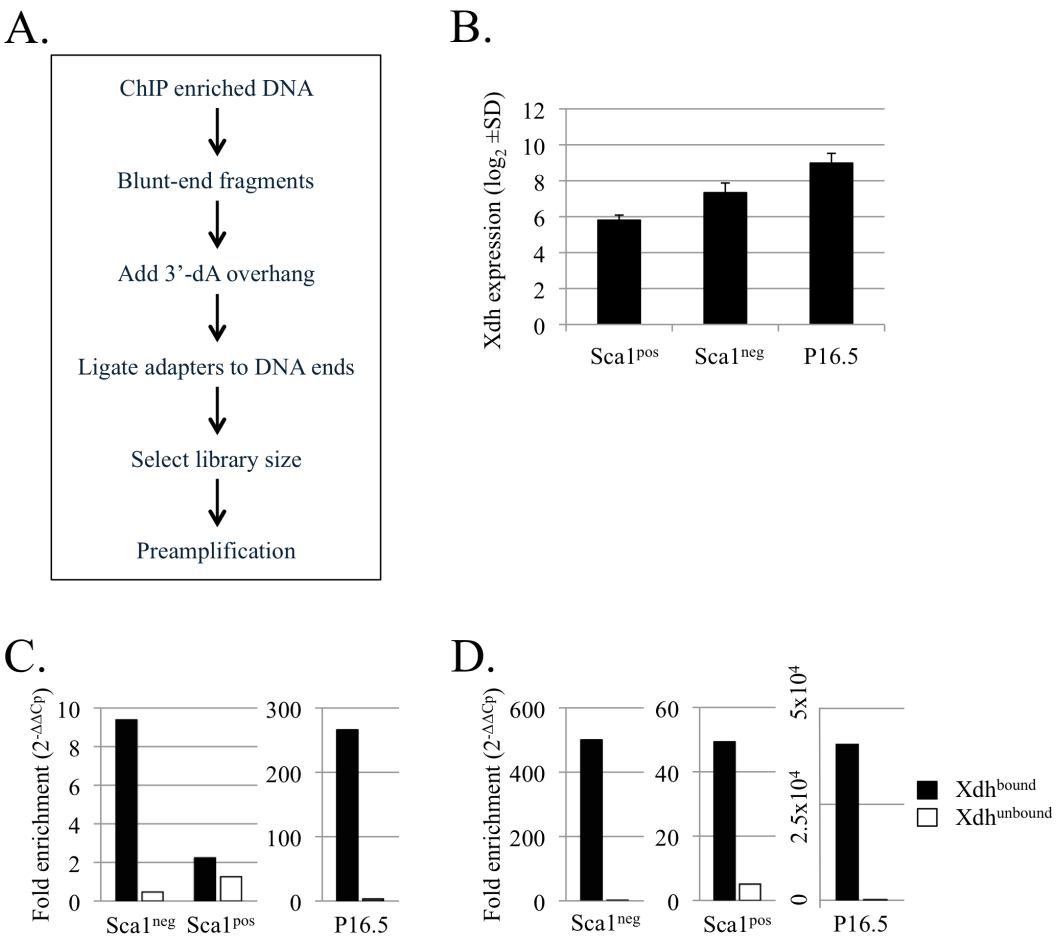
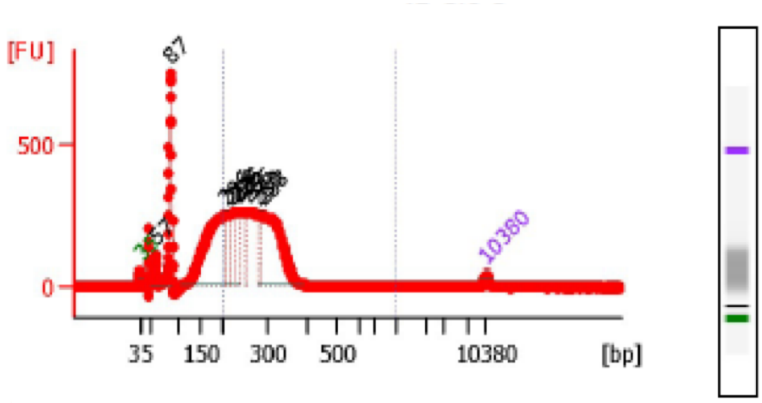


Figure 4.5 DNA sequencing library preparation. (A) General flow chart showing the steps in preparing DNA sequencing libraries using the Illumina ChIP-Seq Sample Preparation kit. **(B)** Average (RMA normalized) expression values (log_2) of *Xdh* from the microarray analysis for $Sca1^{pos}$ (n=3), $Sca1^{neg}$ (n=3) and P16.5 cells (n=4). Values represent the mean, and error bars indicate standard deviations. **(C)** Quantitative PCR assessment of C/EBPβ ChIP using chromatin from sorted $Sca1^{pos}$ and $Sca1^{neg}$ (10-week old virgin females) and from sorted P16.5 alveolar cells (day 16.5 of pregnancy). Chromatin was immunoprecipitated with C19 anti-C/EBPβ or non-immune rabbit IgG. The bar graph represents fold enrichment between input-standardized C/EBPβ and IgG ChIP ($2^{-\Delta\Delta Cp}$). **(D)** Quantitative PCR assessment of DNA sequencing libraries prepared from samples in (C) using the method shown in (A). The bar graph represents fold enrichment between input-standardized C/EBPβ and IgG ChIP ($2^{-\Delta\Delta Cp}$). **Abbreviations:** ChIP, chromatin immunoprecipitation; dA, deoxyadenine; P16.5, pregnant day 16.5 alveolar cells; SD, standard deviation.

A.



B.

Sequencing library	Average size (bp)	Concentration (pg/ul)
Sca1^{neg} C/EBPβ	271	6369.68
Sca1^{neg} IgG	274	12316.96
Sca1^{pos} C/EBPβ	274	11588.24
Sca1^{pos} IgG	287	5613.72
P16.5 C/EBPβ	272	6845.35
P16.5 IgG	278	21610.83

Figure 4.6 Size distribution of DNA sequencing libraries. (A) Representative bioanalyzer trace using an HS CHIP showing the size range and distribution of DNA fragments in the sequencing library. **(B)** Table showing the average fragment length and concentration of each sequencing library submitted for Solexa sequencing.

4.4.3 Analysis of genome-wide C/EBPβ occupancy in three mammary gland cell populations

The ChIP-Seq and microarray data were curated by Dr. Simon Tomlinson using Geneprof, and the workflow is shown in Figure 4.7. The datasets contained between 4-8x10^7 aligned reads, with approximately 60 % of reads uniquely aligning to the mouse genome (Appendix 6). The chromosomal distribution of the reads was similar between the C/EBPβ and control IgG ChIP (Appendix 7). The data presented in Figure 4.8A shows that C/EBPβ binding occurs throughout the genome. Approximately a quarter of C/EBPβ binding occurs in regions between 0-5 kilobase (kb) upstream from the transcription start site (narrow promoter [±500 bp], wide promoter [±2000 bp], 5 kb upstream), a quarter in regions between 10-50 kb upstream of the transcription start site (distal regulatory elements), a quarter in introns, and a quarter of binding sites could not be mapped to known genes. The most highly enriched binding motif found in the C/EBPβ peaks is TTGCGCAA (Fig. 4.8B), which is in agreement with the consensus C/EBP binding motif of RTTGCGYAAY, where R=A/G and Y=C/T (Osada et al., 1996). The number of C/EBPβ binding sites increases from Sca1^{pos} cells (1502) to Sca1^{neg} cells (6976) to P16.5 (14,118) (Fig. 4.8C). As well, almost all of C/EBPβ binding events in Sca1^{pos} cells are present in Sca1^{neg} cells (94.5 %) and P16.5 (96.3 %). Furthermore, most of the binding events in Sca1^{neg} cells are also present in P16.5 (95.8 %), whereas approximately half of the binding events in P16.5 are unique (52.4 %). These observations demonstrate a correlation between the frequency of genome-wide C/EBPβ occupancy and alveolar lineage commitment, as the more committed the cell type to the alveolar lineage, the higher the frequency of C/EBPβ occupancy.

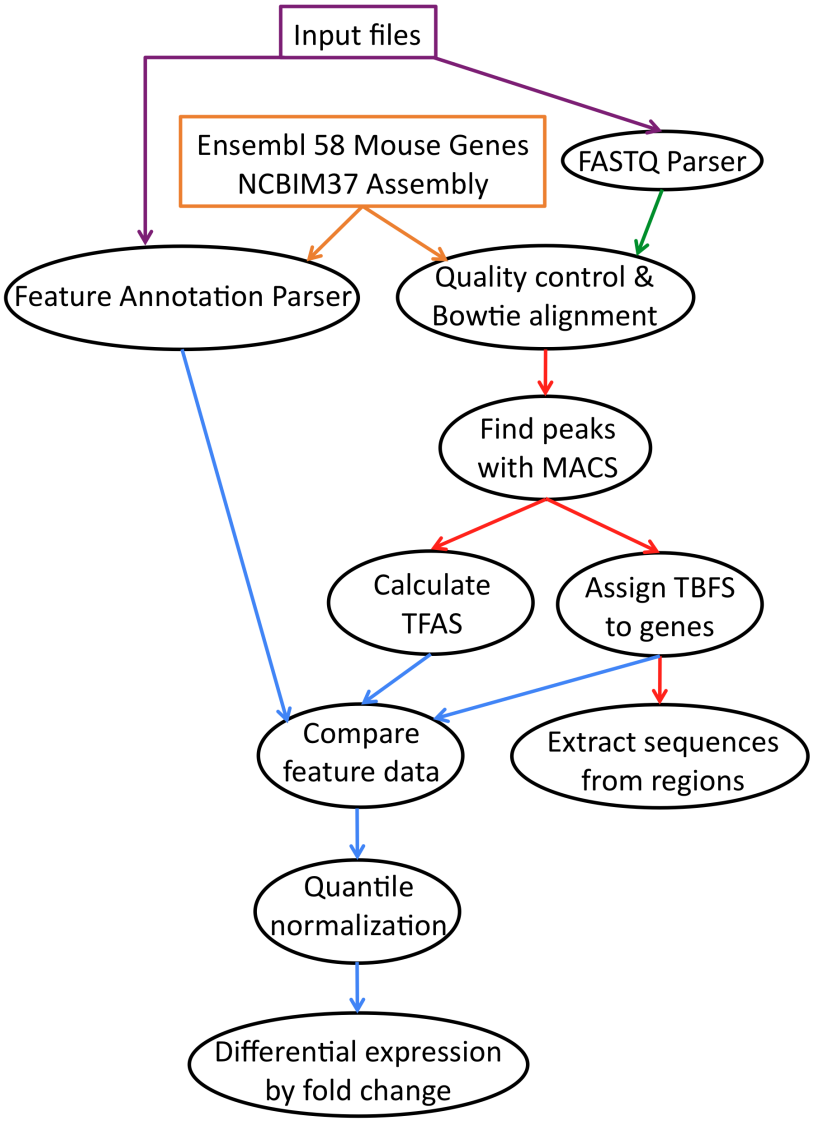


Figure 4.7 Schematic diagram depicting the work flow used in the analysis of the ChIP-Seq data sets. Fifty base pair single-end reads were trimmed to remove adapter sequences and aligned to the mouse genome (Ensembl 58 Mouse genome NCBIM37 Assembly) using Bowtie. For quality control, the mean quality score was calculated for each data set by summing up all of the quality scores for each read and dividing the sum by the length of the read. A mean quality score of ≥8 passed the quality control filter. Peak calling was performed using MACS, and transcription factor binding sites (TFBS) were annotated to genes using the default parameters on Geneprof. To compare C/EBPβ binding with gene expression, the annotated binding peaks were compared to RMA-normalized and quality-filtered expression values from the mouse exon arrays. Supplementary information related to this figure (quality control analyses of ChIP-Seq data sets) is shown in Appendices 6&7. **Abbreviations:** MACS, model-based analysis of ChIP-Seq; TFAS, transcription factor associated score; TFBS, transcription factor binding site.

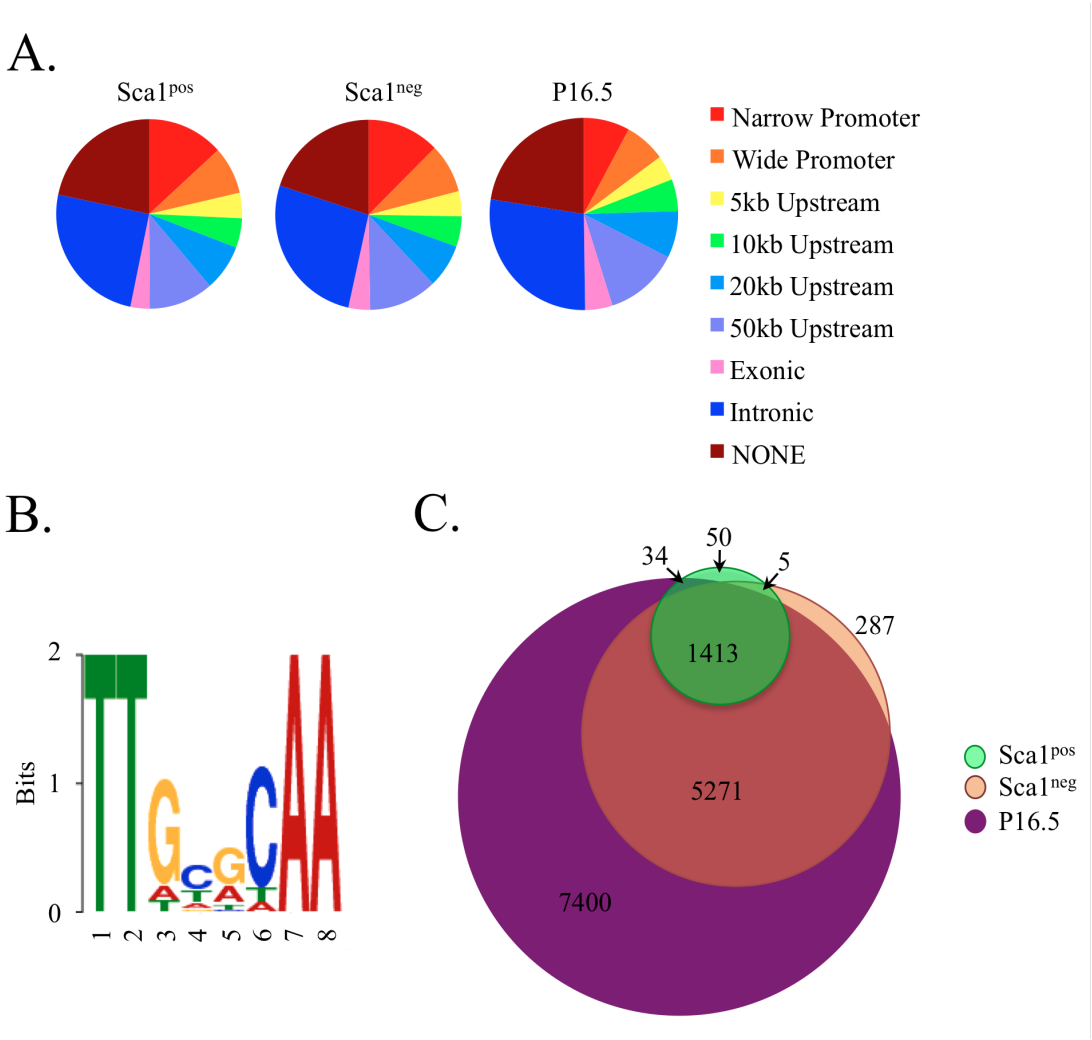


Figure 4.8 Genome-wide binding characteristics of C/EBPβ. (A) C/EBPβ occupancy in different genomic regions in Sca1^{pos} cells, Sca1^{neg} cells and P16.5 alveolar cells. Transcription factor binding sites were annotated to genomic regions based on the distance of the binding peak from known transcription start sites. **(B)** Motif analysis showing that the canonical C/EBPβ binding site is the most highly enriched motif in the binding peaks. **(C)** The Venn diagram depicts the total number of C/EBPβ binding events in Sca1^{pos}, Sca1^{neg} and P16.5 alveolar cells. Almost all Sca1^{pos} binding events are shared with Sca1^{neg} and P16.5, and almost all Sca1^{neg} binding events are shared with P16.5. About one half of P16.5 binding events are unique. **Abbreviations:** P16.5, pregnant day 16.5 alveolar cells; kb, kilobase.

To further investigate how C/EBPβ occupancy correlates with lineage commitment, the pattern of C/EBPβ binding within each gene signature was assessed. The frequency of C/EBPβ occupancy is low within the ductal gene signature in $Sca1^{pos}$ cells, with approximately 7.5 % of ductal-lineage specific genes bound by C/EBPβ (Fig. 4.9). In $Sca1^{neg}$ cells, the frequency of C/EBPβ occupancy within the alveolar-lineage specific gene signature is 43.1 %, and in P16.5 alveolar cells, C/EBPβ binding within the mature alveolar gene signature is 48 %. Functional annotation clustering analysis of the C/EBPβ bound genes within $Sca1^{pos}$ cells did not generate any clusters that met the enrichment score cutoff of ≥2, whereas the unbound genes formed the same functional clusters as described in Chapter 2 (Fig. 4.10A). In $Sca1^{neg}$ cells, one functional annotation cluster passed the enrichment score cutoff of ≥2 for genes not bound by C/EBPβ (Fig. 4.10B). This cluster included the gene ontology (GO) terms: cell motion/migration/localization/motility and receptor tyrosine kinase/protein/cell surface signalling. For C/EBPβ-bound genes, two clusters emerged that contain genes involved in functions related to glycoproteins/signal peptide/disulfide bonding and extracellular matrix interactions (GO terms: carbohydrate/pattern/polysaccharide /glycosaminoglycan binding and cell surface) (Fig. 4.10C). In pregnant alveolar cells, eight clusters emerged for unbound genes (Fig. 4.11A). These genes encode extracellular matrix/basement membrane/collagen components, mitotic factors, proteins involved in microtubule-based processes and GTPase regulators. Genes bound by C/EBPβ in P16.5 are also enriched in mitotic factors, extracellular matrix components and proteins involved in secretion/glycosylation/disulfide bond formation (Fig. 4.11B). In addition to these shared processes, C/EBPβ bound genes are also involved in ATP-binding, cytoskeleton organization, preparation of chromosomes for cell division, and inflammatory responses. These analyses reveal that C/EBPβ target genes have distinct functions from non-targets within the two alveolar cell populations, but not in ductal progenitor cells.

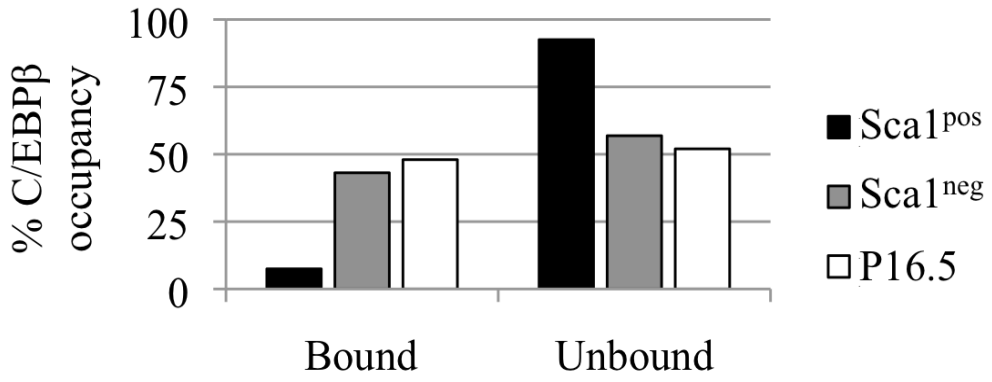
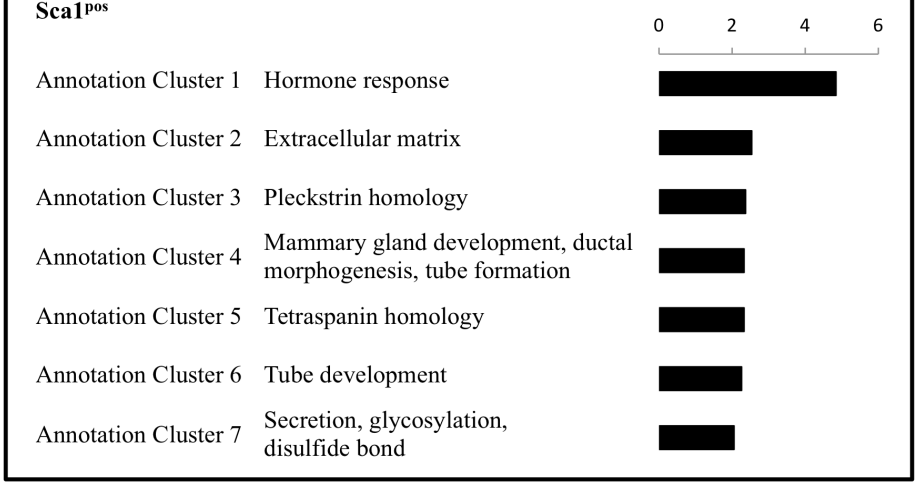


Figure 4.9 C/EBPβ occupancy within the gene signatures of Sca1^{pos}, Sca1^{neg} and P16.5 alveolar cells. Lineage-specific gene lists for Sca1^{pos} and Sca1^{neg} luminal cells were generated by performing a pair-wise comparison between the normalized expression values from the microarrays. Differential expression was considered significant if the false detection rate (FDR) was ≤0.01. The lineage-specific gene list for P16.5 was generated by performing two pair-wise comparisons (P16.5 vs virgin Sca1^{pos} and P16.5 vs virgin Sca1^{neg}) using normalized expression values (differential expression was considered significant in FDR ≤0.01) and extracting the common unique genes from each comparison. Each gene signature was then compared against the annotated C/EBPβ target genes, and the frequency of C/EBPβ binding is shown as a bar graph (bound + unbound = 100 %). **Abbreviations:** P16.5, pregnant day 16.5 alveolar cells.

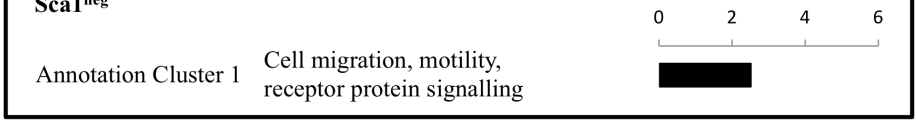
A.

No C/EBPβ binding Sca1^{pos}	Function	Enrichment score
Annotation Cluster 1	Hormone response	
Annotation Cluster 2	Extracellular matrix	
Annotation Cluster 3	Pleckstrin homology	
Annotation Cluster 4	Mammary gland development, ductal morphogenesis, tube formation	
Annotation Cluster 5	Tetraspanin homology	
Annotation Cluster 6	Tube development	
Annotation Cluster 7	Secretion, glycosylation, disulfide bond	



B.

No C/EBPβ binding Sca1^{neg}	Function	Enrichment score
Annotation Cluster 1	Cell migration, motility, receptor protein signalling	



C.

Yes C/EBPβ binding Sca1^{neg}	Function	Enrichment score
Annotation Cluster 1	Glycoprotein, signal peptide, disulfide bond	
Annotation Cluster 2	Extracellular matrix interactions	

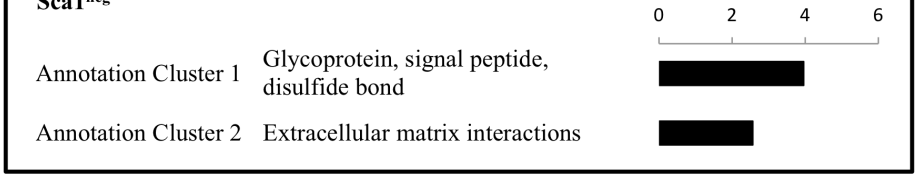
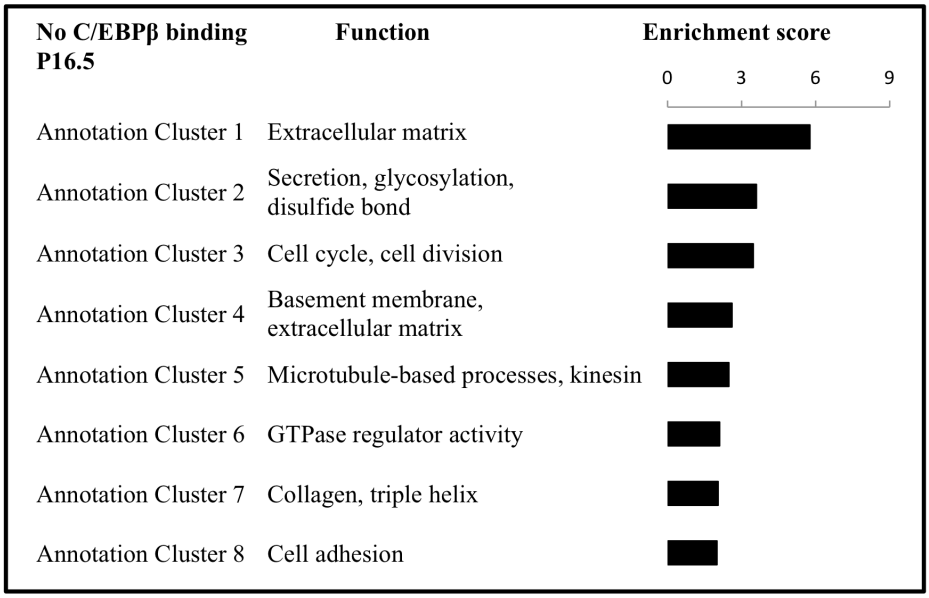


Figure 4.10 Functional annotation clustering analysis of C/EBPβ bound and unbound genes in Sca1^{pos} and Sca1^{neg} luminal cells. (A) Functional clusters were generated from ductal-lineage specific genes. The transcript cluster IDs of differentially expressed genes (from Fig. 2.7) were grouped based on C/EBPβ binding (bound and unbound genes), and these gene lists analyzed using the DAVID Bioinformatics Resource using default parameters. Functional clusters were included in the figure if they had an enrichment score of ≤2. No functional clusters passed this enrichment score cut-off for C/EBPβ-bound genes in Sca1^{pos} cells. **(B&C)** Functional clusters were generated from alveolar-lineage specific genes using the same analysis as in (A). Supplementary information related to this figure (complete list of functional clusters) is shown in Appendix 8-10.

A.



B.

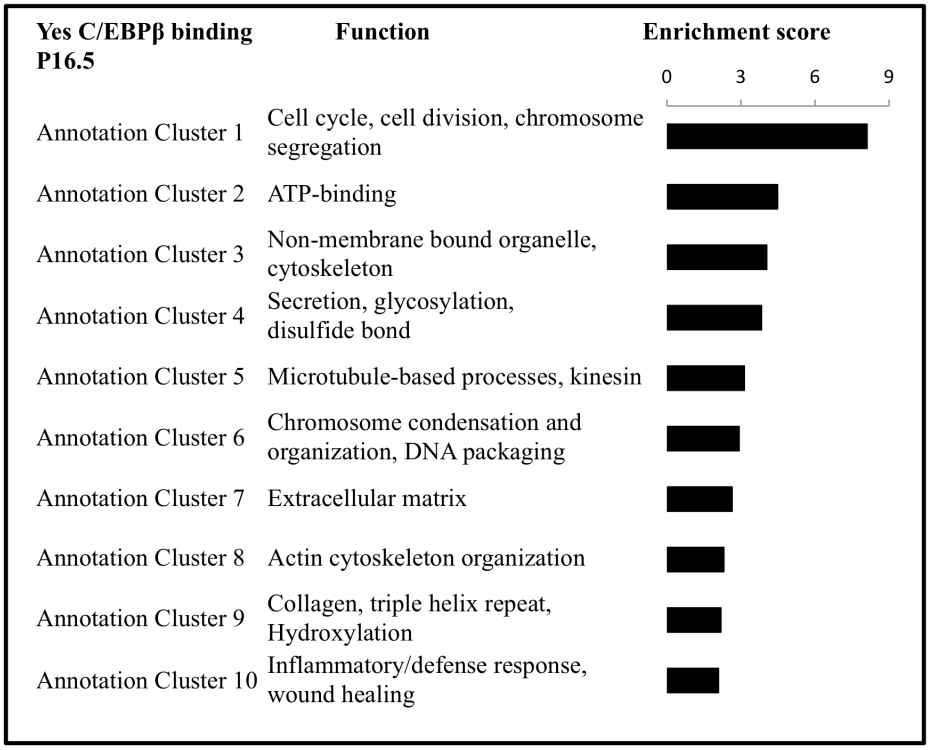


Figure 4.11 Functional annotation clustering analysis of C/EBPβ bound and unbound genes in P16.5 alveolar cells. (A&B) Functional clusters were generated from mature alveolar genes using the same analysis as in Fig. 4.10. Supplementary information related to this figure (complete list of functional clusters) is shown in Appendix 8-10. **Abbreviations:** P16.5, pregnant day 16.5 alveolar cells.

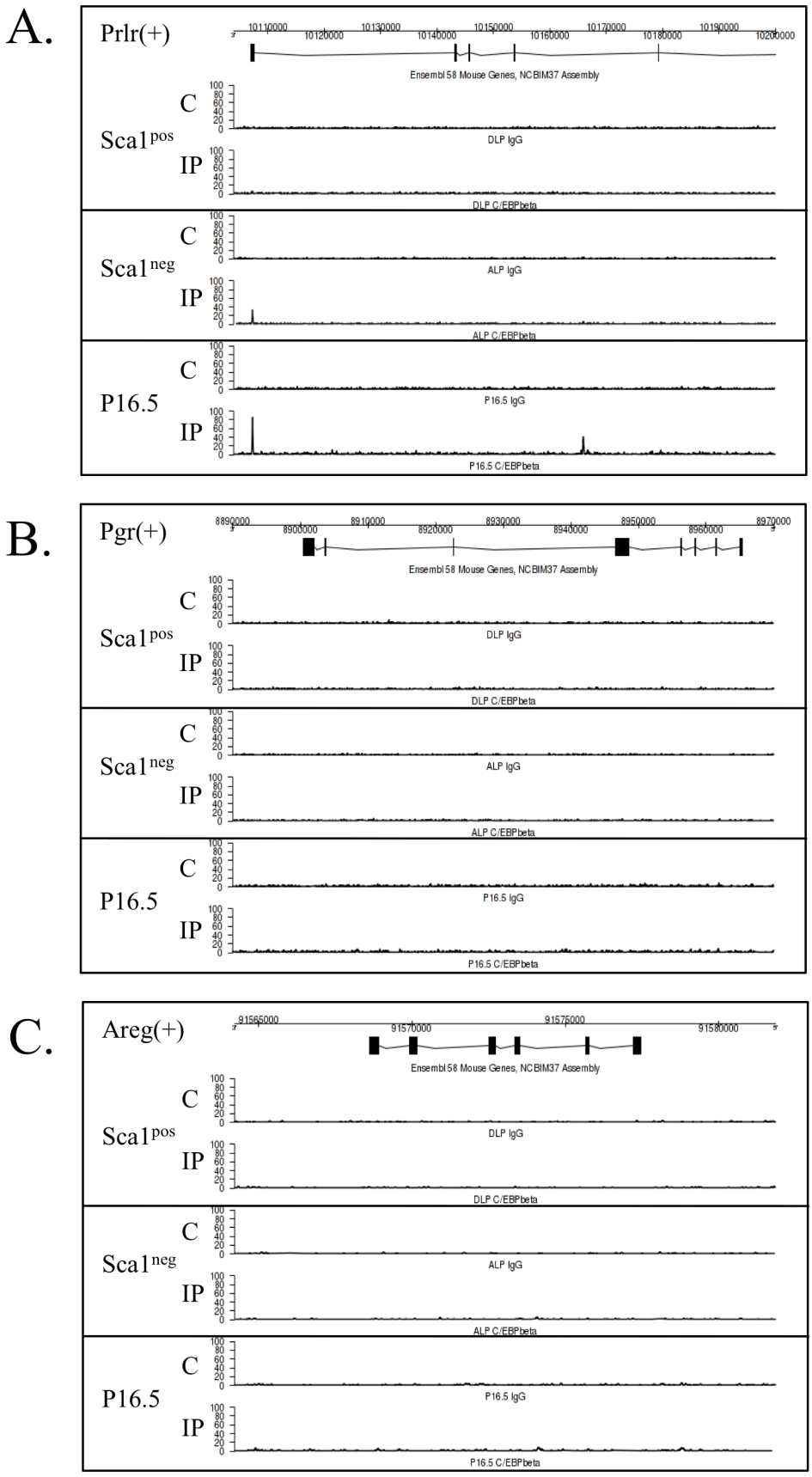
To further demonstrate that C/EBPβ occupancy is important in alveolar development, wiggle tracks were generated for the ductal and alveolar gene signatures shown in Figure 4.1. The ductal genes *Pgr*, *Areg* and *Foxa1* are not bound by C/EBPβ in any of the three cell types (Fig. 4.12B,C,E). Whereas *Prlr* and *Esr1* (Fig. 4.12A,D), both of which are ductal genes in the virgin mammary gland but have known functions in alveolar development during pregnancy and lactation, are bound by C/EBPβ, with increased occupancy (higher peaks=increased number of reads) observed in the progression towards differentiated alveolar cell identity (i.e., lowest/no peak in $Sca1^{pos}$, mid-size peak in $Sca1^{neg}$, large peak in P16.5). On the other hand, all of the alveolar genes from Figure 4.1 (*Csn3*, *Xdh*, *Ltf*, *Elf5*) are bound by C/EBPβ (Fig. 4.12F-I). The same general pattern of binding can be observed for most of the C/EBPβ target genes, whereby small or no binding peaks are seen in $Sca1^{pos}$ cells, medium-sized binding peaks are seen in $Sca1^{neg}$ cells, and large binding peaks are present in P16.5. As well, several of the genes acquire new binding sites in P16.5 alveolar cells (e.g., *Prlr*, *Csn3*, *Xdh*) that are not present in the progenitor cells. These examples, together with the bioinfomatics analysis of global C/EBPβ occupancy in each gene signature (Fig. 4.9) further demonstrate that C/EBPβ preferentially binds alveolar genes, and the level of C/EBPβ occupancy is correlated with the level of differentiation in alveolar cells.

Some of the common luminal genes were also bound by C/EBPβ, although the pattern of binding is more variable. For example, *Krt18* and *Sox9* are only bound in P16.5, with no binding observed in $Sca1^{neg}$ and $Sca1^{pos}$ cells (Fig. 4.12J,K). *Klf5* is bound by C/EBPβ in all three populations, albeit the peak in $Sca1^{pos}$ cells is very small (Fig. 4.12L). There is one small binding peak in $Sca1^{neg}$ cells and two small binding peaks in P16.5 for *Gata3*, albeit not in the proximal promoter region of the gene (Fig. 4.12M). It is possible that these peaks correspond not to *Gata3*, but to another gene located in this genomic region (*930412013Rik*), whose transcription start site is near the position of the two small peaks. Two other distinct peaks are also observed within *Gata3* in $Sca1^{pos}$ cells, but these peaks are present in the IgG control as well, therefore they would not be considered real binding events. *Id2*, a gene demonstrated to be a target of C/EBPβ *in vitro* (using NIH3T3 cells), is not bound in

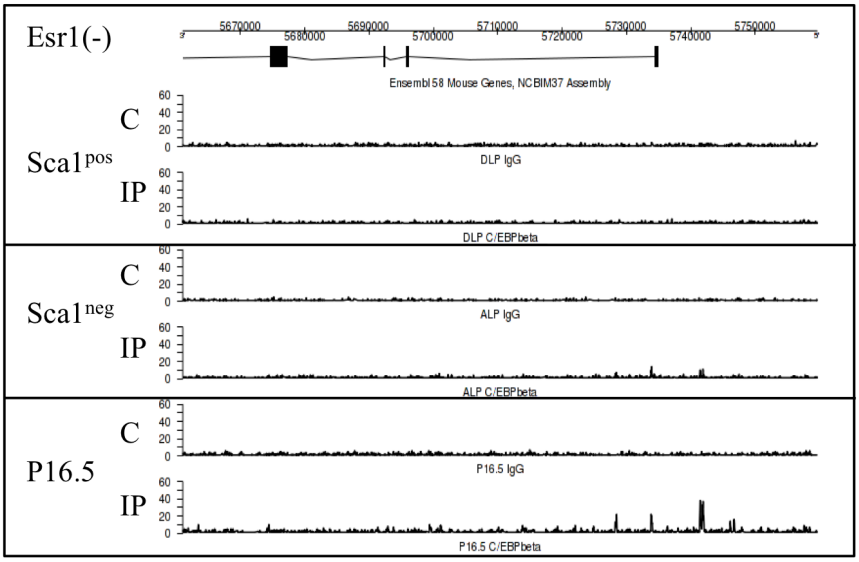
Sca1^{pos}, Sca1^{neg} or P16.5 cells (Fig. 4.12N), consistent with the results obtained from the ChIP experiments presented in Figure 4.3B. *Csn2*, on the other hand, is bound by C/EBPβ in P16.5 (Fig. 4.12O) but is not bound in Sca1^{pos} and Sca1^{neg} cells. The absence of C/EBPβ binding in Sca1^{neg} cells was corroborated by the qPCR assay presented in Figure 4.3B that was initially designed as a negative control for *Csn2* in the region near these binding sites (between exons 3 and 4). The number of genes bound by C/EBPβ in Sca1^{pos} cells is fewer than in Sca1^{neg} cells and P16.5, and only a few unique genes were found that contain a C/EBPβ binding peak. One such example, shown in Figure 4.12P, is the gene encoding the V-type proton ATPase subunit B (*Atp6v1b1*).

Four target genes were chosen thus far for qPCR validation of the Solexa sequencing results. Quantitative PCR assays were designed against the newly identified C/EBPβ binding sites (bound) and another region within each of the genes where binding was not observed (unbound). These qPCR assays were then tested on ChIP-ed DNA from Sca1^{pos}, Sca1^{neg} and P16.5 alveolar cells (Fig. 4.13). These qPCR data corroborate the binding patterns observed in the wiggle tracks. For example, the level of enrichment for *Prlr* is highest in P16.5, then Sca1^{neg}, and lowest in Sca1^{pos}. This same pattern is observed for *Csn3* and *Csn2*, while the opposite pattern is observed for the ductal-specific binding of *Atp6v1b1*, consistent with the wiggle track in Figure 4.12P.

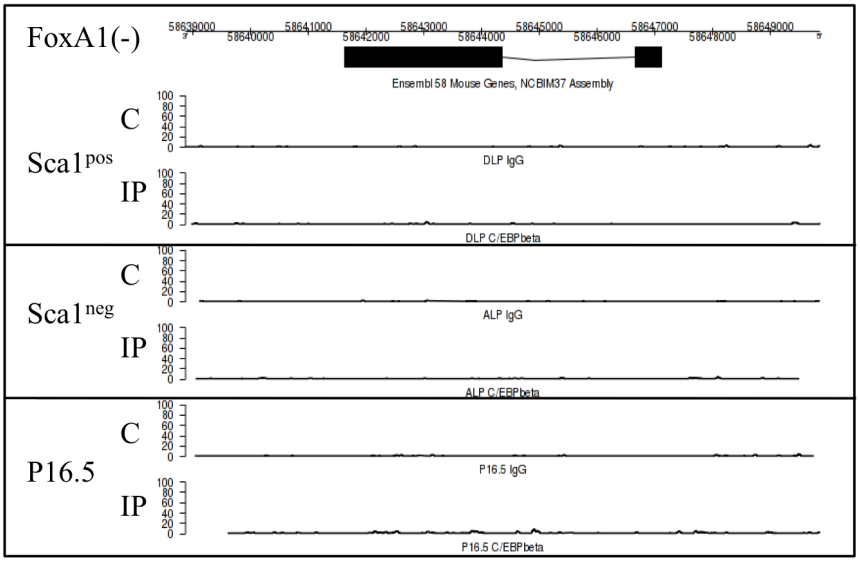
Overall, these data demonstrate that C/EBPβ binds alveolar genes, and C/EBPβ occupancy in these genes increases as the cells progress towards alveolar differentiation. These observations support the hypothesis that C/EBPβ directs alveolar lineage specification and commitment. The role of C/EBPβ in the ductal lineage is equivocal, as almost all of the genes bound by C/EBPβ in the ducts are shared with Sca1^{neg} cells and P16.5, and querying the C/EBPβ-bound genes in the ductal gene signature did not identify a role in any unique cellular functions.


A.
Prlr(+)
10110000
10120000
10130000
10140000
10150000
10160000
10170000
10180000
10190000
10200000
Ensembl 58 Mouse Genes, NCBIM37 Assembly
C
Sca1pos
IP
DLP IgG
DLP C/EBPbeta
C
Sca1neg
IP
ALP IgG
ALP C/EBPbeta
C
P16.5
IP
P16.5 IgG
P16.5 C/EBPbeta
B.
Pgr(+)
8890000
8900000
8910000
8920000
8930000
8940000
8950000
8960000
8970000
Ensembl 58 Mouse Genes, NCBIM37 Assembly
C
Sca1pos
IP
DLP IgG
DLP C/EBPbeta
C
Sca1neg
IP
ALP IgG
ALP C/EBPbeta
C
P16.5
IP
P16.5 IgG
P16.5 C/EBPbeta
C.
Areg(+)
91565000
91570000
91575000
91580000
Ensembl 58 Mouse Genes, NCBIM37 Assembly
C
Sca1pos
IP
DLP IgG
DLP C/EBPbeta
C
Sca1neg
IP
ALP IgG
ALP C/EBPbeta
C
P16.5
IP
P16.5 IgG
P16.5 C/EBPbeta

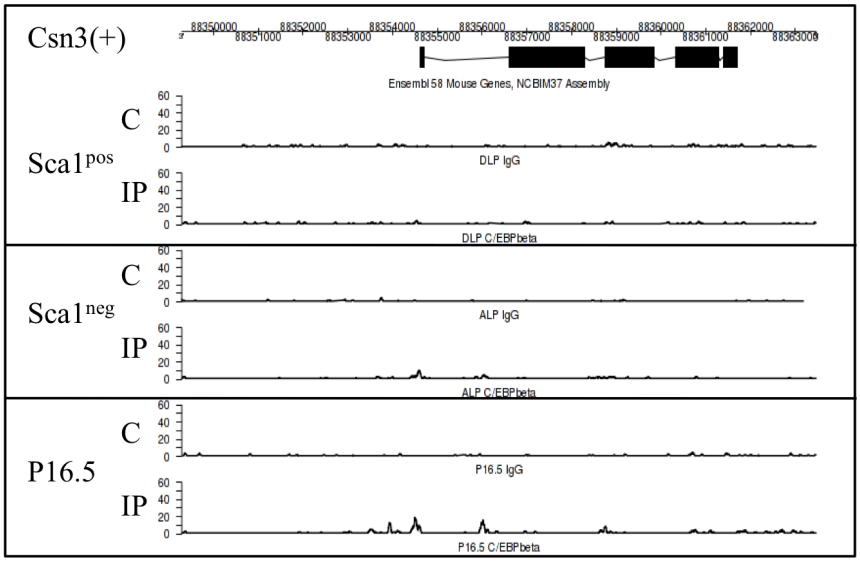
D.


Esr1(-)
5670000
5680000
5690000
5700000
5710000
5720000
5730000
5740000
5750000
Ensembl 58 Mouse Genes, NCBIM37 Assembly
C
Sca1pos
IP
DLP IgG
DLP C/EBPbeta
C
Sca1neg
IP
ALP IgG
ALP C/EBPbeta
C
P16.5
IP
P16.5 IgG
P16.5 C/EBPbeta

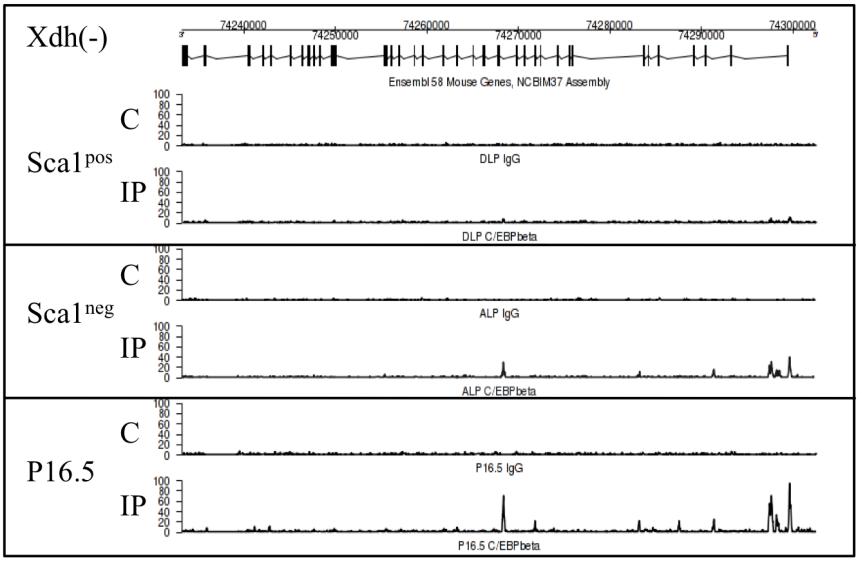
E.


FoxA1(-)
58639000
58640000
58641000
58642000
58643000
58644000
58645000
58646000
58647000
58648000
58649000
Ensembl 58 Mouse Genes, NCBIM37 Assembly
C
Sca1pos
IP
DLP IgG
DLP C/EBPbeta
C
Sca1neg
IP
ALP IgG
ALP C/EBPbeta
C
P16.5
IP
P16.5 IgG
P16.5 C/EBPbeta

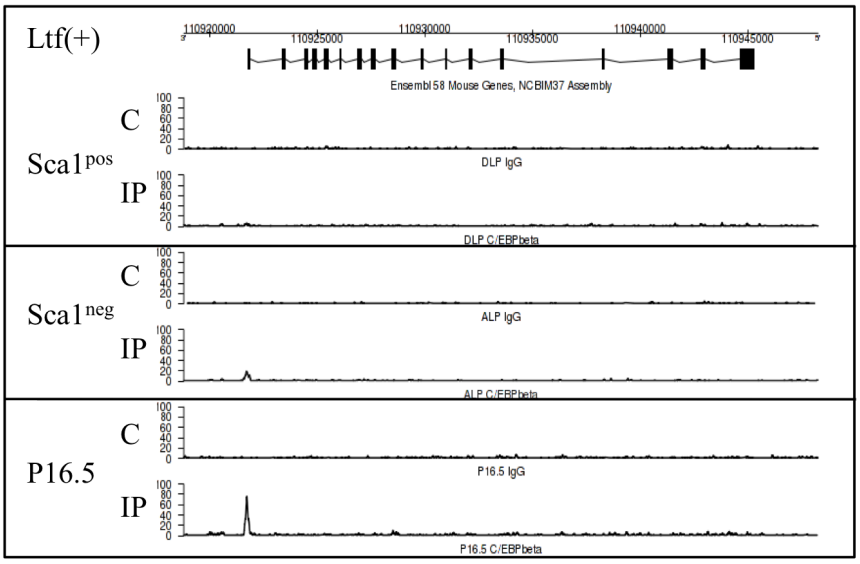
F.


Csn3(+)
88350000
88351000
88352000
88353000
88354000
88355000
88356000
88357000
88358000
88359000
88360000
88361000
88362000
88363000
Ensembl 58 Mouse Genes, NCBIM37 Assembly
C
Sca1pos
IP
DLP IgG
DLP C/EBPbeta
C
Sca1neg
IP
ALP IgG
ALP C/EBPbeta
C
P16.5
IP
P16.5 IgG
P16.5 C/EBPbeta

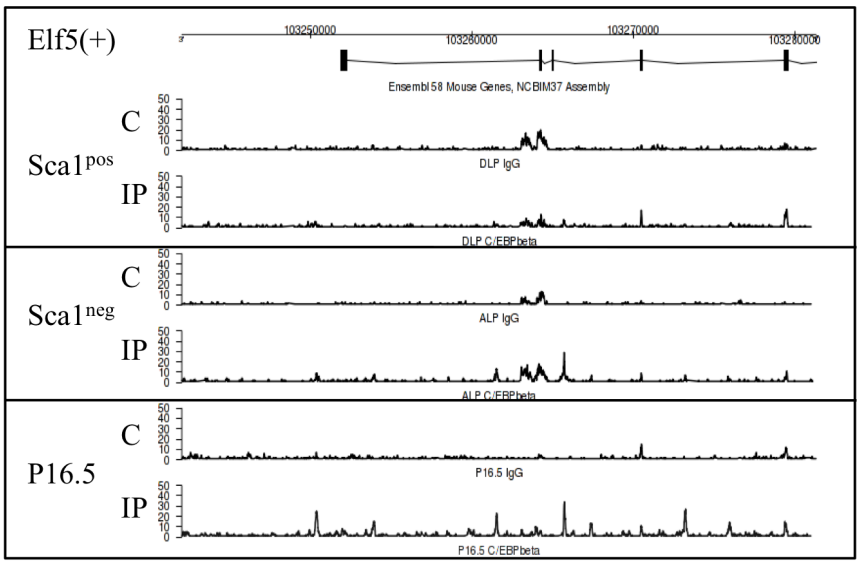
G.


Xdh(-)
74240000
74250000
74260000
74270000
74280000
74290000
74300000
Ensembl 58 Mouse Genes, NCBIM37 Assembly
Sca1pos
C
DLP IgG
IP
DLP C/EBPbeta
Sca1neg
C
ALP IgG
IP
ALP C/EBPbeta
P16.5
C
P16.5 IgG
IP
P16.5 C/EBPbeta

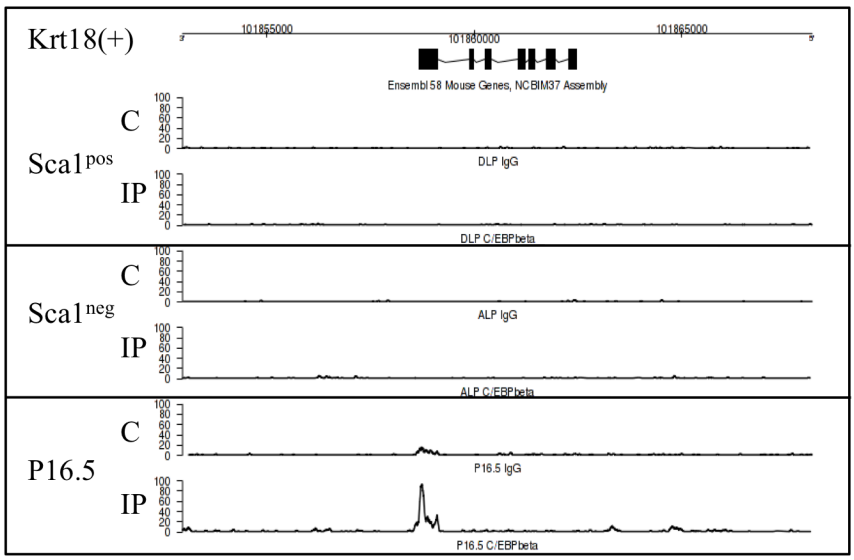
H.


Ltf(+)
110920000
110925000
110930000
110935000
110940000
110945000
Ensembl 58 Mouse Genes, NCBIM37 Assembly
Sca1pos
C
DLP IgG
IP
DLP C/EBPbeta
Sca1neg
C
ALP IgG
IP
ALP C/EBPbeta
P16.5
C
P16.5 IgG
IP
P16.5 C/EBPbeta

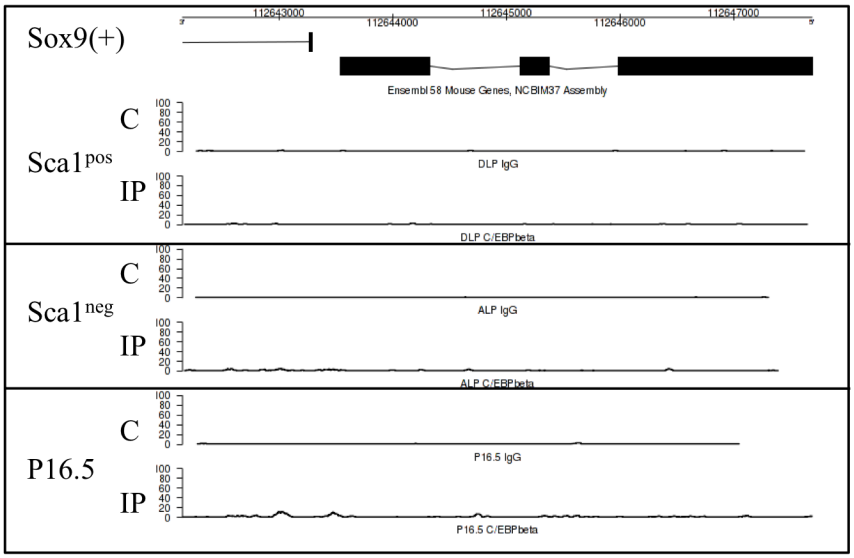
I.


Elf5(+)
103250000
103260000
103270000
103280000
Ensembl 58 Mouse Genes, NCBIM37 Assembly
Sca1pos
C
DLP IgG
IP
DLP C/EBPbeta
Sca1neg
C
ALP IgG
IP
ALP C/EBPbeta
P16.5
C
P16.5 IgG
IP
P16.5 C/EBPbeta

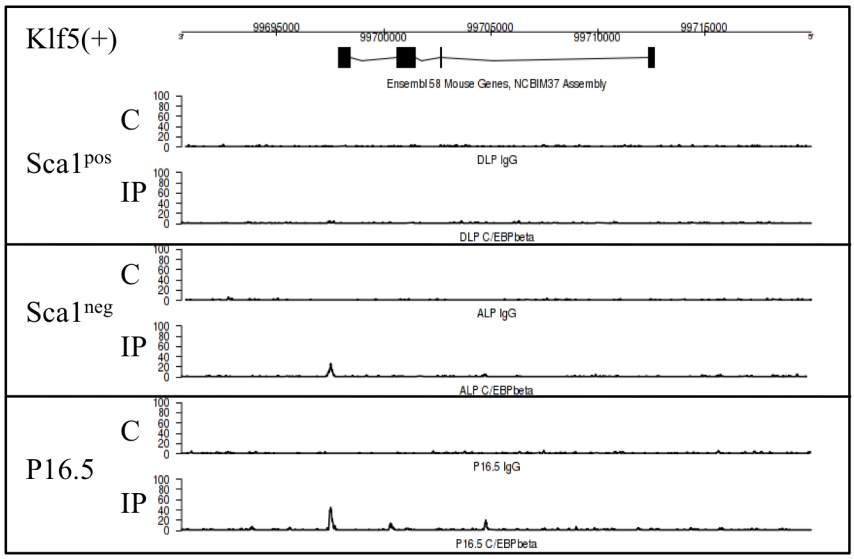
J.


Krt18(+)
101855000
101860000
101865000
Ensembl 58 Mouse Genes, NCBIM37 Assembly
C
Sca1pos
IP
DLP IgG
DLP C/EBPbeta
C
Sca1neg
IP
ALP IgG
ALP C/EBPbeta
C
P16.5
IP
P16.5 IgG
P16.5 C/EBPbeta

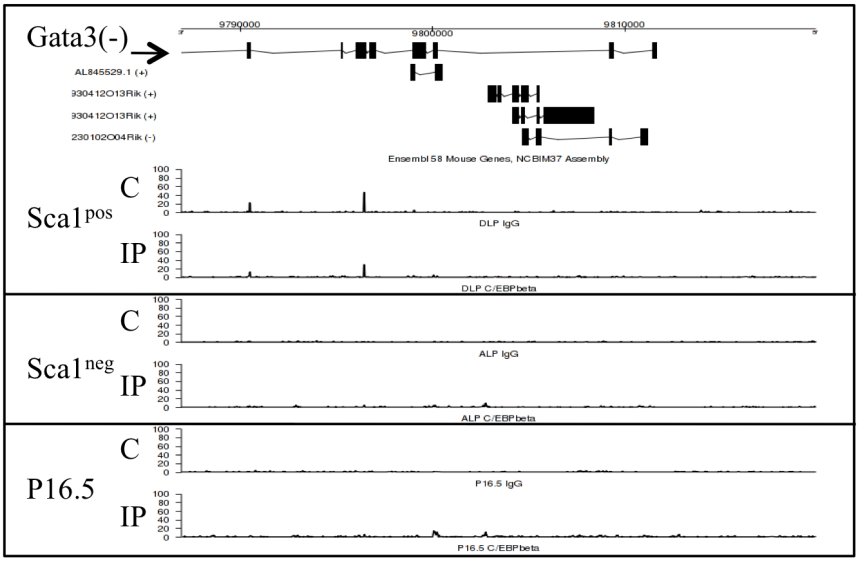
K.


Sox9(+)
112643000
112644000
112645000
112646000
112647000
Ensembl 58 Mouse Genes, NCBIM37 Assembly
C
Sca1pos
IP
DLP IgG
DLP C/EBPbeta
C
Sca1neg
IP
ALP IgG
ALP C/EBPbeta
C
P16.5
IP
P16.5 IgG
P16.5 C/EBPbeta

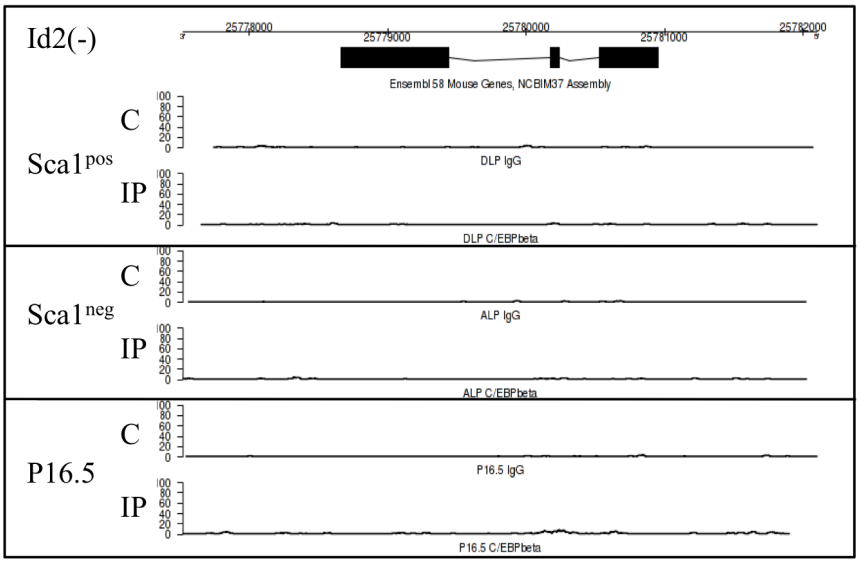
L.


Klf5(+)
99695000
99700000
99705000
99710000
99715000
Ensembl 58 Mouse Genes, NCBIM37 Assembly
C
Sca1pos
IP
DLP IgG
DLP C/EBPbeta
C
Sca1neg
IP
ALP IgG
ALP C/EBPbeta
C
P16.5
IP
P16.5 IgG
P16.5 C/EBPbeta

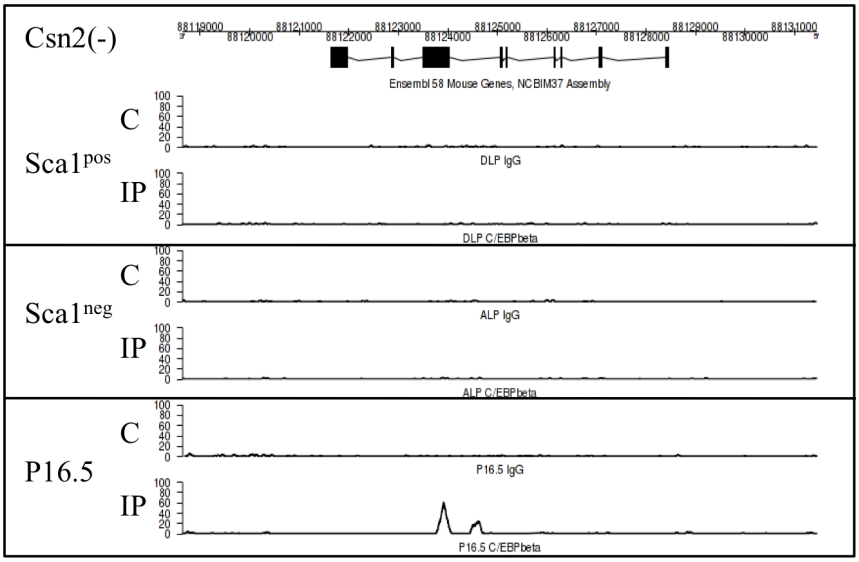
M.


Gata3(-)
9790000
9800000
9810000
AL845529.1 (+)
330412O13Rik (+)
330412O13Rik (+)
230102O04Rik (-)
Ensembl 58 Mouse Genes, NCBIM37 Assembly
C
Sca1pos
IP
DLP IgG
DLP C/EBPbeta
C
Sca1neg
IP
ALP IgG
ALP C/EBPbeta
C
P16.5
IP
P16.5 IgG
P16.5 C/EBPbeta

N.


Id2(-)
25778000
25779000
25780000
25781000
25782000
Ensembl 58 Mouse Genes, NCBIM37 Assembly
C
Sca1pos
IP
DLP IgG
DLP C/EBPbeta
C
Sca1neg
IP
ALP IgG
ALP C/EBPbeta
C
P16.5
IP
P16.5 IgG
P16.5 C/EBPbeta

O.


Csn2(-)
88118000
88119000
88120000
88121000
88122000
88123000
88124000
88125000
88126000
88127000
88128000
88129000
88130000
88131000
Ensembl 58 Mouse Genes, NCBIM37 Assembly
C
Sca1pos
IP
DLP IgG
DLP C/EBPbeta
C
Sca1neg
IP
ALP IgG
ALP C/EBPbeta
C
P16.5
IP
P16.5 IgG
P16.5 C/EBPbeta

P.

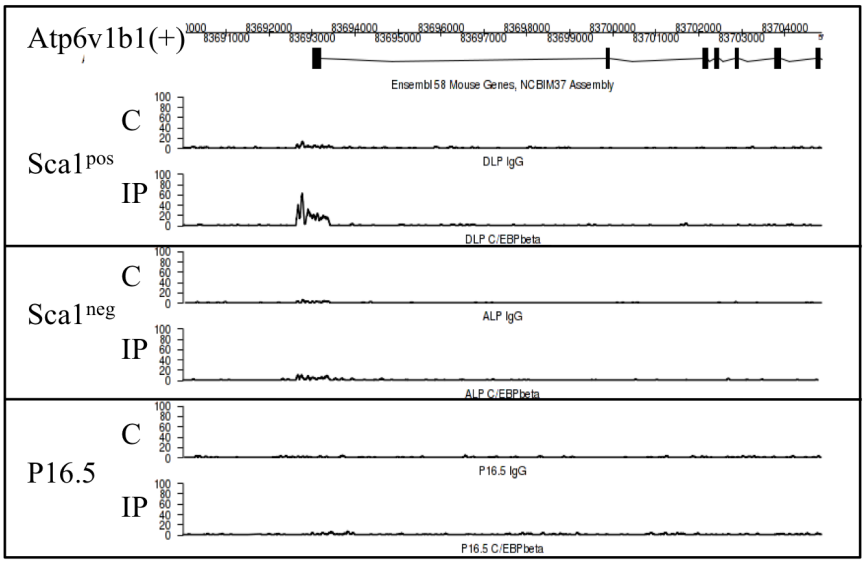


Figure 4.12 Examples of C/EBPβ binding in Sca1^{pos}, Sca1^{neg} and P16.5 alveolar cells. Wiggle tracks were generated using Geneprof to illustrate the pattern of C/EBPβ binding in ductal-lineage and alveolar-lineage genes (from the genes used as examples in Fig. 2.7). A schematic diagram for each gene is shown at the top (to scale), followed by wiggle tracks of the ChIP-Seq data for non-immune IgG (C=control) and C/EBPβ (IP=immunoprecipitation) for Sca1^{pos}, Sca1^{neg} and P16.5 alveolar cells. The orientation for each gene is denoted by a plus (+) or minus (-) sign after the gene name, where plus (+) indicates forward orientation, and minus (-) indicates reverse orientation. **(A-E)** Ductal lineage genes. **(F-I)** Alveolar lineage genes. **(J-M)** Common luminal genes. (N-O) Published C/EBPβ target genes. **(P)** Example of a gene that is bound by C/EBPβ only in Sca1^{pos} cells. **Abbreviations:** C, IgG immunoprecipitation; IP, C/EBPβ immunoprecipitation; P16.5, pregnant day 16.5 alveolar cells.

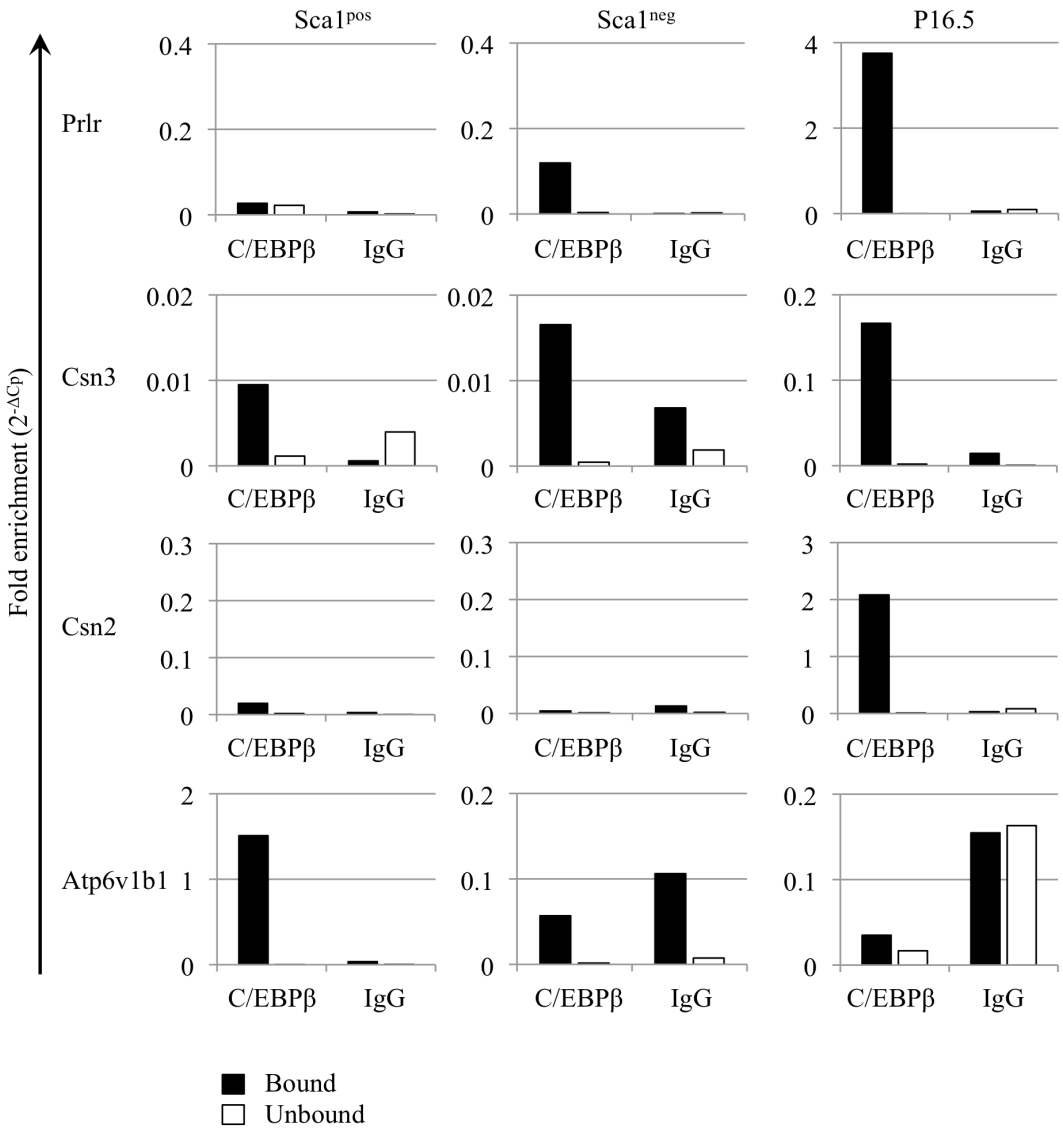


Figure 4.13 Validation of C/EBPβ binding. Quantitative PCR assays were designed against newly identified binding sites and intralocus negative controls for four C/EBPβ target genes (*Prlr*, *Csn3*, *Csn2*, *Atp6vb1*). Enrichment for C/EBPβ-binding was assessed using DNA sequencing libraries prepared with the Illumina ChIP-Seq Sample Preparation kit for $Sca1^{pos}$, $Sca1^{neg}$ and P16.5 alveolar cells. The bar graphs represent fold enrichment ($2^{-\Delta Cp}$) after standardization against non-immunoprecipitated gDNA (input).

4.4.4 Elf5 is a direct C/EBPβ target gene in vivo

The observations that *Elf5* expression in down-regulated in *Cebpb*-/- mice (Fig. 3.8), and that *Elf5* is a direct target of C/EBPβ *in vivo* (Fig. 4.12I), led us to propose that C/EBPβ may regulate alveolar priming and differentiation, at least in part, by direct activation of *Elf5* expression. The *Elf5* gene is transcribed from two distinct promoters (Fig. 4.14A). The wiggle track in Figure 4.12I and the follow-up ChIP experiment shown in Figure 4.14A demonstrate that C/EBPβ binds the downstream proximal promoter (variant 1, TSS1) of *Elf5*, with no enrichment observed at the upstream promoter (variant 2, TSS2). To demonstrate that C/EBPβ binding activates Elf5 transcription, *in vitro* luciferase reporter assays were performed. The promoter region of the *Elf5* variant 1 containing the C/EBPβ binding site was cloned into a luciferase reporter vector and co-transfected with an expression vector encoding C/EBPβ (pcDNA3-*Cebpb*). Four-fold higher *Elf5* promoter-driven luciferase activity was measured with co-transfection of 100 ng of the C/EBPβ expression vector compared to the empty vector control (Fig. 4.14B), demonstrating that C/EBPβ binding is sufficient to activate *Elf5* gene transcription.

To further investigate the relationship between C/EBPβ and ELF5 downstream signalling, C/EBPβ binding was assessed in several known ELF5 target genes. *Ccnd2*, *Muc4* and *Igfbp4* were recently identified as direct *in vivo* targets of ELF5 in unfractionated pregnant (P17.5) mammary tissue (Escamilla-Hernandez et al., 2010). C/EBPβ binding was found in *Ccnd2* and *Muc4* in P16.5 alveolar cells, but not in $Sca1^{neg}$ or $Sca1^{pos}$ cells (Fig. 4.15A-B). *Igfbp4* was also found to be a C/EBPβ target in P16.5 cells and $Sca1^{neg}$ cells, but not in $Sca1^{pos}$ cells (Fig.4.16A). These data suggest that C/EBPβ and ELF5 may co-regulate the expression some target genes. However, due to the limited number of known Elf5 target genes in the mammary gland, these investigations were not exhaustive. To address this shortcoming, the ChIP method was optimized for ELF5 to globally identify its target genes in $Sca1^{pos}$, $Sca1^{neg}$ and P16.5 cells. A preliminary ELF5 ChIP experiment is shown in Figure 4.16B. Using two commercially available anti-ELF5 antibodies (H62 and AP7394C), a high level of enrichment is observed in the bound promoter of *Ifgbp4*, whereas no

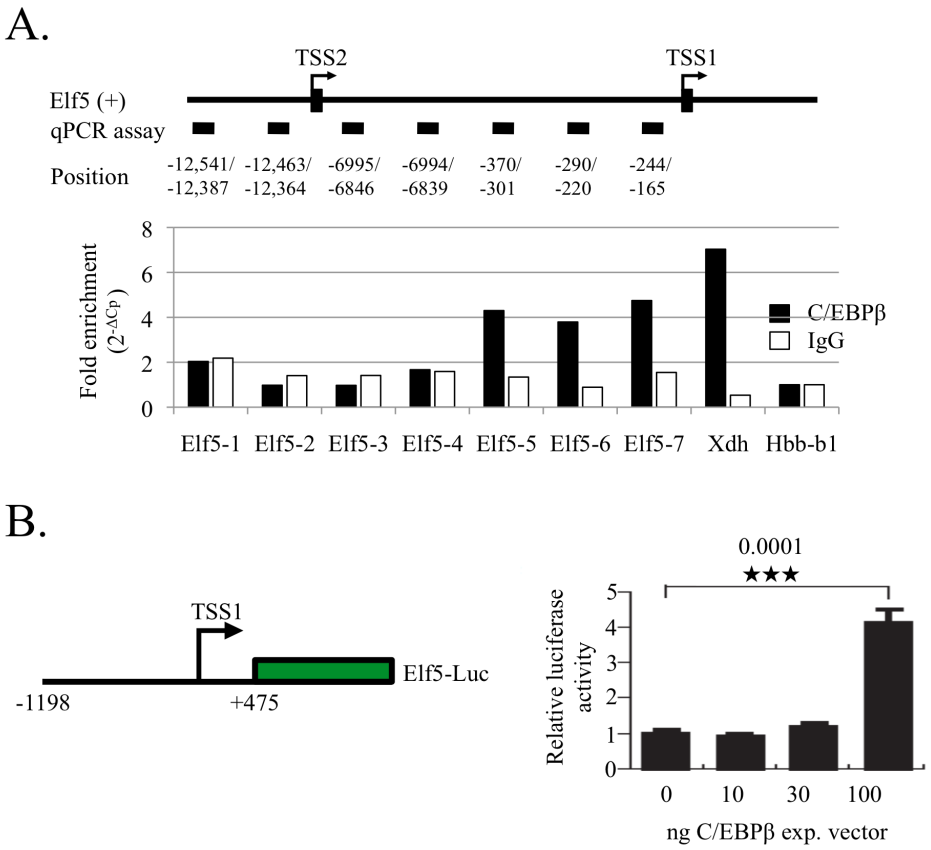


Figure 4.14 Elf5 is a C/EBPβ target gene. (A) Schematic diagram (not to scale) showing the alternative first exons of the *Elf5* gene, giving rise to variant 1 (TSS1) and variant 2 (TSS2). Both transcripts encode the same protein. Underneath the schematic diagram, the position of the seven amplicons used for ChIP analysis is indicated by black boxes/numbers. Numbering is relative to TSS1, where TSS1=1. The graph below show a representative ChIP experiment performed on sorted Sca1^{neg} cells from 10-week old virgin females. Chromatin was immunoprecipitated with C19 anti-C/EBPβ or non-immune rabbit IgG. The values represent fold enrichment ($2^{-\Delta Cp}$) after standardization against non-immunoprecipitated gDNA (input) and normalization against *Hbb-b1* (encoding β-globin). Xdh was used as a positive control. **(B)** Schematic diagram (left) showing the design of the firefly luciferase reporter vector using the promoter region of *Elf5* variant 1. The numbering in the schematic diagram is relative to TSS1. The firefly luciferase reporter construct was co-transfected with pcDNA3-Cebpb expression vector and phRL-CMV (renilla luciferase internal control), and promoter activation was assessed using the DLR kit (Promega). The values in the bar graph represent the mean firefly-to-renilla ratio from 3 experimental wells, each read in triplicate, after background subtraction (empty vector signal), and setting the baseline value to 1. Error bars indicate standard deviation between replicate wells. The p-value was calculated using an unpaired, one-tailed Student's T-test, and significance is shown as a star and value above the graph. **Abbreviations:** TSS, transcription start site; Luc, firefly luciferase.

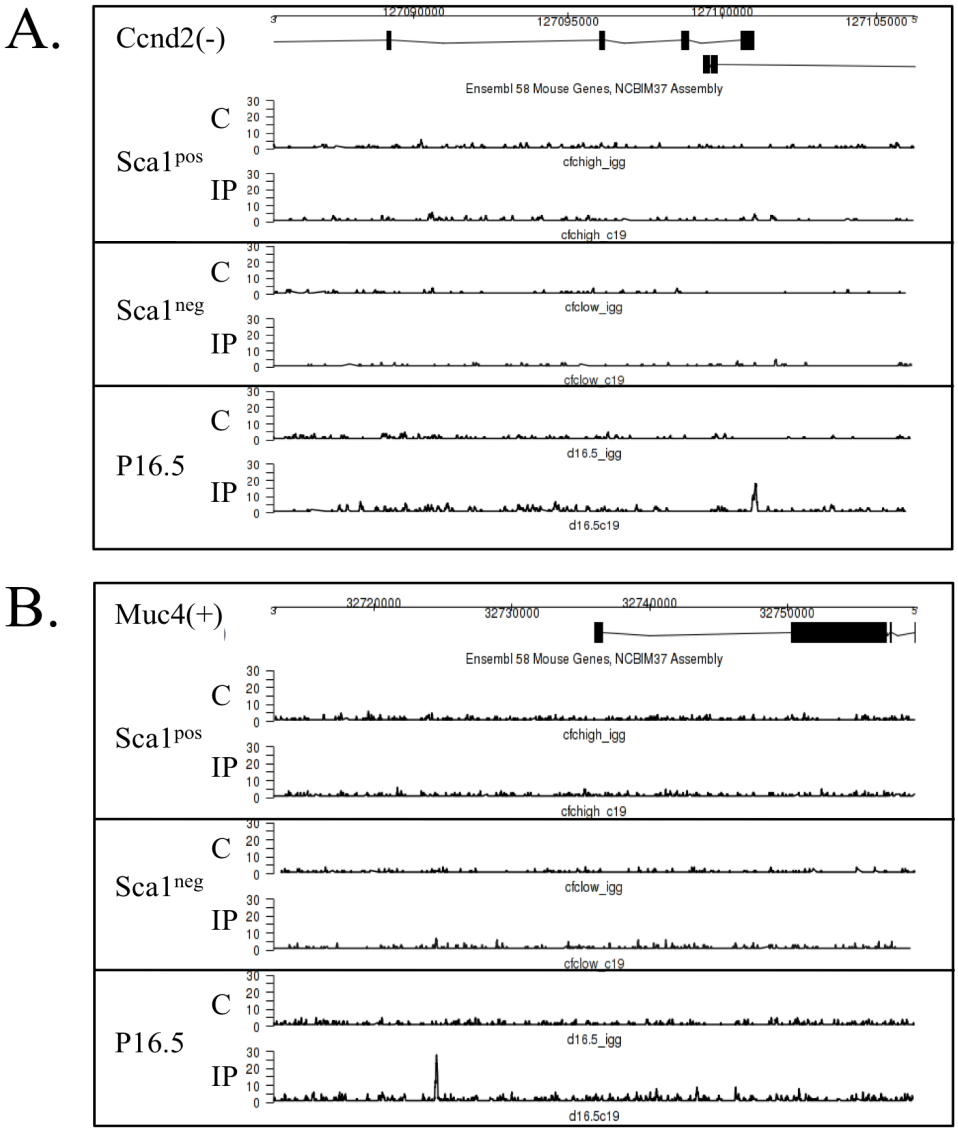


Figure 4.15 Co-regulation of *Ccnd2* and *Muc4* by C/EBPβ and ELF5. (A&B) Wiggle tracks were generated using Geneprof to illustrate the pattern of C/EBPβ binding in two published ELF5 target genes. A schematic diagram for each gene is shown at the top (to scale), followed by wiggle tracks of the ChIP-Seq data for non-immune IgG (C=control) and C/EBPβ (IP=immunoprecipitation) for Sca1^{pos}, Sca1^{neg} and P16.5 alveolar cells. The orientation for each gene is denoted by a plus (+) or minus (-) sign after the gene name, where plus (+) indicates forward orientation, and minus (-) indicates reverse orientation.

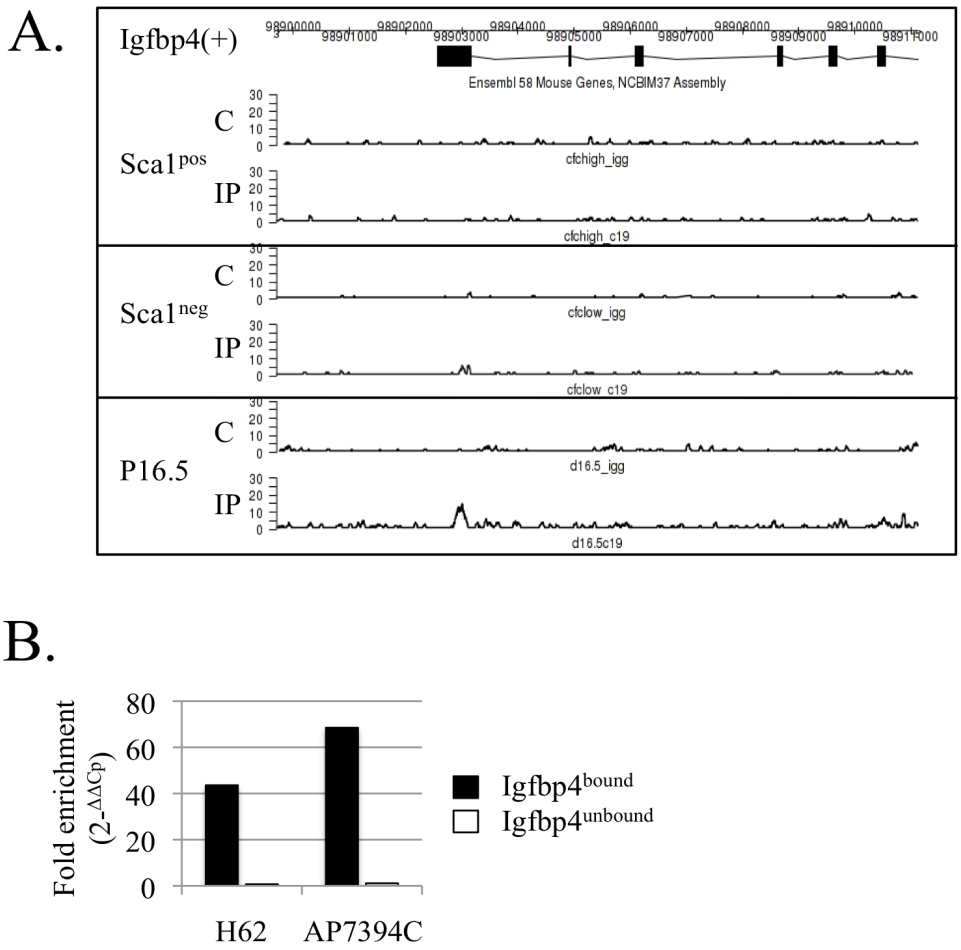


Figure 4.16 Co-regulation of *Igfbp4* by C/EBPβ and ELF5. (A) The wiggle track was generated using Geneprof to illustrate the pattern of C/EBPβ binding in *Igfbp4*, a published ELF5 target gene. A schematic diagram is shown at the top (to scale), followed by wiggle tracks of the ChIP-Seq data for non-immune IgG (C=control) and C/EBPβ (IP=immunoprecipitation) for $Sca1^{pos}$, $Sca1^{neg}$ and P16.5 alveolar cells. The orientation for *Igfbp4* is denoted by a plus (+), and indicates forward orientation. **(B)** Preliminary ELF5 ChIP experiment performed on sorted alveolar cells from day 1 of lactation. Chromatin was immunoprecipitated with anti-ELF5 antibodies (H62 and AP7394C) or non-immune rabbit IgG. The values represent fold enrichment between input-standardized ELF5 and IgG ChIP ($2^{-\Delta\Delta Cp}$).

enrichment is observed in the unbound intralocus control. ChIP-Seq experiments to assess genome-wide ELF5 binding are currently ongoing and will be continued in my next post in Prof. Nerlov's laboratory.

4.5 Discussion

4.5.1 Comparison of gene signatures in Sca1^{pos}, Sca1^{neg} and P16.5 cells

While mature ductal cells are present in the mammary gland around puberty, alveolar cells only mature during mid to late pregnancy and become fully differentiated secretory cells following parturition (Richert et al., 2000; Anderson et al., 2007). Therefore, to assess the molecular mechanisms through which C/EBPβ regulates alveolar development, a second time point in late pregnancy (P16.5) was chosen. Choosing this time point in late pregnancy rather than after birth (when alveoli are fully differentiated) was based on the observations that C/EBPβ is most highly expressed during pregnancy, with low levels of expression observed in the virgin and lactating mammary gland (Gigliotti and DeWille, 1998; Robinson et al., 1998; Sabatakos et al., 1998).

To investigate C/EBPβ function in P16.5 luminal cells, we first analyzed their global gene expression profile. The flow cytometric strategy used to isolate P16.5 cells is similar to that used to isolate luminal progenitors in the adult virgin gland, except that singlets were selected for based on the FACSJAZZ-specific parameters trigger pulse width and SSC-488 rather than FCS-A and FCS-H, as on the FACSAriaII (Fig. 4.1A). While the intention was to isolate Sca1^{pos} and Sca1^{neg} luminal cells from the pregnant mammary gland, the Sca1^{pos} cells were almost completely absent (Fig. 4.1A). During early pregnancy, alveolar progenitor cells rapidly proliferate, and during mid to late pregnancy, they begin to differentiate (Anderson et al., 2007). These *in vivo* morphological and functional changes are consistent with the alteration in the cellular composition of the luminal compartment (compare Fig. 2.5A and Fig 4.1A), which shows that by day 16.5 of pregnancy, almost all of the luminal cells

have a Sca1^{neg} phenotype (Fig. 4.1A), which we have previously shown to be associated with an alveolar cell identity (Chapter 2).

P16.5 luminal cells differentially express 1198 unique genes when compared to Sca1^{neg} and Sca1^{pos} cells (FDR ≤0.01). The expression of alveolar genes is highly upregulated in P16.5 (e.g., *Csn3*, *Xdh*), while the expression of ductal genes (e.g., *Prg*, *Areg*) is further suppressed (Fig. 4.1B,C). *Prlr* and *Esr1* are exceptions to this pattern, as they are more highly expressed in P16.5 than in Sca1^{neg} cells. However, this is consistent with their established biological role (discussed in Chapter 1) in alveolar differentiation during pregnancy (Brisken et al., 1999; Mallepell et al., 2006). Interestingly, the expression of the alveolar gene *Ltf* is down-regulated in P16.5 cells when compared with Sca1^{pos} and Sca1^{neg} cells (Fig. 4.1B,C). Lactoferrin is an iron-binding protein found in milk, with known functions in innate immunity (Clarkson et al., 2004). In unfractionated mammary tissue, *Ltf* is expressed at low levels in the virgin and pregnant gland, with a sharp, but moderate increase at parturition, and a substantial increase during involution (Molenaar et al., 1996; Clarkson et al., 2004). Its role as a soluble innate defense factor is consistent with an increase in *Ltf* expression during periods when the mammary gland is most vulnerable to infections (lactation) or when immune cells are known to be active (involution) (Clarkson et al., 2004). In reference to *Ltf* expression in ewes and cows, Molenaar noted that overall expression appears to be inversely related to the degree of alveolar differentiation (i.e., secretory alveolar cells have lower *Ltf* expression than pre-secretory alveolar cells) (Molenaar et al., 1996). However, these findings are inconsistent with Clarkson's observation showing that *Ltf* expression is approximately 2-fold higher in lactating mammary glands than during pregnancy (Clarkson et al., 2004). Our gene expression arrays (n=3 for Sca1^{neg} and Sca1^{pos} cells, n=4 for P16.5) show that *Ltf* expression is higher in Sca1^{neg} and Sca1^{pos} cells than in P16.5, but we did not assess *Ltf* expression in the lactating mammary gland, therefore the relationship between our findings and the published studies is not clear (Fig. 4.1C). At present, the functional significance of our observation that *Ltf* expression is lower in P16.5 mammary glands than in virgin progenitor cells is unclear, and provides an interesting area for future investigation.

In summary, these data show that P16.5 luminal cells are enriched in genes required for alveolar function, which is consistent with the biological function of the mammary gland at this developmental time point. As well, together with the analyses of the virgin mammary gland described in Chapter 2, these data confirm the hypothesis that the alveolar cell identity is contained within the $Sca1^{neg}$ luminal cell compartment.

4.5.2 C/EBPβ ChIP-Seq

Due to the lack of known C/EBPβ target genes in the mammary gland, and the need to identify unique gene targets in $Sca1^{pos}$, $Sca1^{neg}$ and P16.5 alveolar cells, a ChIP method was developed and optimized for use on low numbers of primary mouse mammary cells. When these experiments were started in 2008, most ChIP protocols were optimized for use on high numbers of cultured cells ($\geq 10^6$ cells). One ChIP protocol assessing epigenetic modifications from sorted hematopoietic stem cells was available for low cells numbers (10,000 cells/ChIP) (Attema et al., 2007). However, after extensive testing and optimization (not included in this thesis), I was unsuccessful in adapting this protocol to my experiments with lower occupancy transcription factor binding in mammary gland progenitors. Therefore, I set out to develop a ChIP-Seq method based on a ChIP-ChIP protocol developed for use with Drosophila embryos (Sandmann et al., 2007), where systemic optimization was performed for each step including: cross-linking, sonication, preclearing, antibody binding, washing conditions and DNA elution steps. The overall aim was to develop a method that would allow us to perform ChIP-Seq from primary mammary gland cells that could be sorted from a small number of adult mice.

An overview of the standard steps in ChIP are shown in Figure 4.2A. First, optimization of the sonication condition was undertaken, where the concentration of sodium dodecyl sulfate (SDS) (Appendix 5) and the required number of sonication cycles were determined (data not shown). Testing of 0.1 % and 1 % SDS concentrations for sonication was based Dr. Pablo Navarro work, who noted that

some transcription factor antibodies did not work when chromatin is sonicated in the presence of 1 % SDS (unpublished observations). However, for the anti-C/EBPβ antibody C19, 1 % SDS in the sonication buffer is optimal (Appendix 5), and the number of cycles required to obtain DNA fragments between 200-400 bp, which is the optimal size range for Solexa sequencing (Schmidt et al., 2009), is 50 cycles of 30 seconds on/30 seconds off on a high setting using a Bioruptor Diagenode waterbath sonicator (Fig. 4.2B). Next, the conditions for antibody binding and capture were optimized using the published C/EBPβ binding sites in the *Xdh* promoter. The quantity of chromatin for each ChIP was chosen based on two considerations. First, previous experience with Attema's method of using 10,000 cells/ChIP was unsuccessful; therefore, to improve the possibility of a positive result, the number of cells/ChIP was increased. Second, it was important to limit the amount of cells/ChIP to a number that could be sorted from one or a few adult mice (consistent with NC3R recommendations). The number of $Sca1^{neg}$ and $Sca1^{pos}$ cells that can be sorted from one adult mouse varies between experiments, but on average, it is between $1\text{-}2x10^5$ cells/mouse. Therefore, I chose 100,000 cells/ChIP as a starting point for the optimization of the ChIP method. The amount of C/EBPβ antibody added to each ChIP was tested, as well as the effects of preclearing the chromatin prior to ChIP with protein A beads, and the use of different combinations of washing buffers after the ChIP. These experiments are summarized in Appendix 5, and show that preclearing the chromatin prior to antibody binding improves the signal and reduces the background, whereas increased washing leads to loss of antibody binding, as shown by the reduced signal in the Xdh^{bound} qPCR assay (Appendix 5). The amount of antibody used for each immunoprecipitation was also optimized, and 0.5 μg C19/100,000 cells was chosen based on the level of enrichment in the bound compared to the unbound region in *Xdh*, as well as based on the reproducibility of the results between biological replicates from independent experiments (Fig. 4.4). Finally, reverse cross-linking was optimized by comparing two methods and determining the yield of DNA from each (Attema et al., 2007; Sandmann et al., 2007). The first method involves incubating each ChIP sample with SDS and proteinase K overnight at 37 ^oC with moderate agitation, followed by a 6-hour incubation at 65 ^oC. The DNA is then purified using phenol chloroform, and

precipitated in the presence of glycogen (Sandmann et al., 2007). The second method involved incubating each ChIP sample with SDS and proteinase K at 68 ^{o}C for 2 hours with vigorous shaking, followed by DNA purification using phenol chloroform and precipitation in the presence of linear acrylamide and glycogen (Attema et al., 2007). The quantity of DNA obtained from the two methods is comparable (data not shown), therefore the second (shorter) method was chosen for the de-crosslinking of C/EBPβ-bound DNA.

The optimized ChIP method was then used to generate samples for the preparation of DNA sequencing libraries, which were verified by qPCR and the DNA quality was assessed by Bioanalyzer analysis prior to Solexa sequencing (Fig. 4.6A). The ChIP-Seq data were curated by Dr. Simon Tomlinson using Geneprof, a software program developed by Florian Halrbitter, a PhD Candidate in Dr. Tomlinson's laboratory (Halbritter et al., 2012).

4.5.3 Genome-wide C/EBPβ occupancy correlates with alveolar cell fate

Most studies investigating C/EBPβ binding have focused primarily on promoter-proximal regions (Karaya et al., 2005; Seymour et al., 2006; Wang et al., 2012). However, our analysis of the genome-wide binding of C/EBPβ reveals that less than ¼ of binding events occurs within 5 kb of known transcription start sites, and more than 50 % map to previously unknown genomic regions (Fig. 4.8A). These findings are consistent with the observations by Schmidt and Wilson, who characterized the genome-wide occupancy of another C/EBP family member, C/EBPα, in liver tissue, and showed that less than ¼ of the C/EBPα binding events localized to within 3 kb of known transcription start sites (Schmidt et al., 2010). In their study, the role of distal binding events was not investigated, although the authors noted that the binding events closest to the TSS were the most highly conserved between the five vertebrate species studied, and were highly enriched in genes known to be up-regulated in liver organogenesis, suggesting that the most proximal binding events may be the most functionally relevant (Schmidt et al., 2010). Further investigations into the characteristics of distal and proximal C/EBPβ binding sites will provide pertinent

information about the complex mechanisms underlying C/EBPβ action in the mammary gland. Unlike C/EBPα, whose function is conceptually clear-cut (inhibits proliferation, promotes differentiation, suppresses tumorigenesis), C/EBPβ function is more complex and highly context-dependent (Nerlov, 2007). For example, C/EBPβ is an important anti-proliferative agent in promyelocytic leukemia cells, where it promotes differentiation towards the granulocytic lineage (Duprez et al., 2003). On the other hand, it has pro-proliferative properties in keratinocytes, where it is required for RAS-mediated tumourigenesis (Zhu et al., 2002). Furthermore, *Cebpb*-/- females have proliferation and differentiation defects in the mammary gland (Robinson et al., 1998; Seagroves et al., 1998), whereas over-expression of C/EBPβ in a human mammary epithelial cells line (MCF10A) leads to hyper-proliferation and induces epithelial to mesenchymal transition, a hallmark of cancer (Bundy and Sealy, 2003). C/EBPβ has also been linked to human epithelial cancers *in vivo*, and was shown to be an essential effector of cyclin D1-driven mammary tumour formation (Lamb et al., 2003). In mice, over-expression of LIP, the predominant C/EBPβ isoform expressed during pregnancy, led to the development of alveolar hyperplasia, which was subsequently attributed to the cells re-entering S-phase (Zahnow et al., 2001). These studies demonstrate that the *in vivo* C/EBPβ target genes have critical functions in regulating the cell cycle, and de-repression or inhibition of these genes by LIP during pregnancy promotes alveolar cell proliferation. This hypothesis is also supported by the observation that genome-wide C/EBPβ occupancy increases as cells progress towards the differentiated alveolar cell fate (Fig. 4.8C, Fig. 4.9). Few unique binding events are observed in the ductal cell lineage, whereas an increasing number of unique binding events are seen in alveolar progenitor cells and differentiating alveolar cells (P16.5) (Fig. 4.8C), particularly in genes that are alveolar-specific (e.g., *Csn3*, *Xdh*, Fig. 4.12). As shown in Figure 4.9, the overall frequency of C/EBPβ binding in the regulatory region of alveolar lineage-specific genes is significantly higher than that of ductal lineage-specific genes. The acquisition of new binding sites within genes is also likely involved in the specification and proliferation of alveolar cells (Fig. 4.12).

4.5.4 C/EBPβ targets genes perform distinct functions in alveolar cells

Functional annotation clustering revealed that Sca1^{pos} genes bound by C/EBPβ do not have a clear role in ductal morphogenesis. However, these preliminary analyses require further investigation, as it is probable that some of these unique target genes have roles in ductal development that were previously unrecognized. Ongoing analyses not yet available to present in this thesis are assessing the functional role of unbound genes in Sca1^{pos} cells that acquire C/EBPβ binding in the alveolar cell types. These genes are particularly interesting, because they may provide the first direct mechanism for the development of the alveolar lineage from ductal-primed luminal progenitor cells, and would demonstrate that this process is directly controlled by C/EBPβ. If empirical evidence supports this hypothesis, a model could be generated where Sca1^{pos} cells are positioned in the cellular hierarchy upstream of Sca1^{neg} cells, and therefore would be considered common luminal progenitor cells that, upon C/EBPβ binding (thereby inhibiting ductal genes and/or activating alveolar genes), differentiate into Sca1^{neg} and alveolar cells, and in the absence of C/EBPβ binding, would generate mature ductal cells.

Further to these analyses, the genes that display C/EBPβ binding in the alveolar lineage gene signature (Sca1^{neg}) are generally involved in extracellular matrix-associated functions (Fig. 4.10C). These include MET (hepatocyte growth factor receptor), whose function was discussed in Chapter 2, and CD44, a type I transmembrane adhesion receptor that mediates interactions between mammary epithelial cells, as well as between epithelial cells and stroma (Louderbough et al., 2011). CD44 expression has been associated with cancer stem cells, and prospective isolation and transplantation of CD24^{lo}CD44^{pos} cells has shown that this population is highly enriched in tumour forming activity (Al-Hajj et al., 2003). However, contrary to this, the involvement of CD44 in reducing the progression of mammary carcinomas was also shown using a mouse model of spontaneously metastasizing breast cancer (MMTV-PyV mT) (Lopez et al., 2005). These studies demonstrated that loss of CD44 promotes tumour metastasis, and that the interaction of CD44 and its ligand hyaluronan, expressed on stromal cells, restricts tumour invasion (Lopez et

al., 2005). Therefore, the role of CD44 is cancer formation requires further investigation. ESR1^{neg} luminal progenitor cells have been proposed to be the initiating cell type for basal-like breast cancer through MET-induced de-differentiation (Gastaldi et al., 2012), and my work has shown that *Met* is a marker of Sca1^{neg} cells (Chapter 2). *Cd44* expression is also enriched in Sca1^{neg} cells, and both *Cd44* and *Met* are direct C/EBPβ targets in these cells. Interestingly, the expression of *Cd44* is comparable in Sca1^{neg} luminal cells and MaSCs, and approximately 4- and 10-fold higher than in Sca1^{pos} cells and P16.5, respectively (data not shown). The expression of *Cd44* in myoepithelial cells has previously been observed, and its role in mediating the organization of the luminal and myoepithelial cells layer was demonstrated in the *Cd44*-/- mouse (Louderbough et al., 2011). *Cd44*-/- mice have impaired ductal elongation and TEB formation during puberty, as well as disrupted organization of the two epithelial cell layers, as demonstrated by the intermixing of luminal and myoepithelial cells (Louderbough et al., 2011). The biological importance of C/EBPβ-mediated regulation of *Cd44* expression in Sca1^{neg} cells is intriguing. The role of *Met* and *Cd44* in tumorigenesis, and the observation that they are directly regulated by C/EBPβ in Sca1^{neg} cells, should prompt further investigations into the potentially directive role of C/EBPβ in the development of basal-like breast cancers. As well, these observations corroborate the findings that ESR1^{neg} luminal progenitor cells (i.e., Sca1^{neg} cells) are the most likely cells-of-origin that generate basal-like breast tumours (Gastaldi et al., 2012).

The genes not bound by C/EBPβ in Sca1^{neg} cells are involved with cell migration/motility and receptor protein signalling, and include the two chemokine ligands *Cxcl2* and *Cxcl3* (Fig. 4.10B). CXCLs are secreted chemokines that attract neutrophils, and CXCL3 was shown to be a normal component of milk (Rainard et al., 2008). However, the biological reason for the constitutively high expression of *Cxcl3* in the mammary gland in the absence of inflammation remains unclear (Rainard et al., 2008). Another gene found in this cluster is the forkhead box transcription factor FOXC1. Recently, the role of FOXC1 in the formation of basal-like breast cancer was demonstrated (Tkocz et al., 2011). These studies showed that co-operative repression of *Foxc1* expression by GATA3 and BRCA1 is required for

normal breast differentiation, and failure to repress *Foxc1* expression contributes to drug resistance and increased tumour aggressiveness in *in vitro* models of basal-like breast cancer (Tkocz et al., 2011). Ectopic FOXC1 expression can also induce de-differentiation of mature mammary epithelial cells into a progenitor-like state (Bloushtain-Qimron et al., 2008). Like CXCL3, the function of FOXC1 during normal mammary gland development is unknown. As these examples demonstrate, most of the information available for many of the lineage-specific genes identified are related to cancer formation, and little is known about their biological function during normal mammary gland development.

During pregnancy, the mammary gland undergoes significant structural changes that result not only from the proliferation and differentiation of alveolar cells, but also from restructuring of extracellular matrix interactions between epithelial cells, as well as between the epithelium and stroma, to enable the gland to transform into a secretory organ. The genes involved in these processes are clearly reflected in the functional annotation clusters for both C/EBPβ bound and unbound genes in P16.5 alveolar cells (Fig. 4.11). The most highly enriched clusters are involved in mitosis, and a significant portion of C/EBPβ-mediated genes are directly involved in the preparation and stabilization of chromosomes for cell division, such as *Ncaph* (chromosome condensation), *Nusap1* (microtubule assembly at metaphase), *Cenpa* (formation of centromere) and *Sgol1* (stabilization of centromeres). C/EBPβ also regulates genes involved in inflammation, defense response and wound healing, such as members of the complement pathway (e.g., *C1qb*, *C3*, *C5ar1*, *Cd55*), *Lat* (T-cell antigen receptor signalling), and *Ptafr* (platelet activating factor receptor). The binding of C/EBPβ to genes involved in immunity is not surprising, as its role as an essential regulator of acute phase proteins in the liver in response to inflammation is well described (Poli, 1998). However, C/EBPβ-mediated expression of immune genes in other tissues (e.g., fat, brain, ovaries, mammary gland) suggests that it may have a general role in regulating immune-response genes that is not related to inflammation.

The role for C/EBPβ-regulated expression of immune molecules in the mammary gland may have evolutionary origins. Vorbach posits that the mammary gland developed from mucosal epithelium, whose primary function was to protect the animal from pathogens (Vorbach et al., 2006). Evidence to support this hypothesis stems from the observation that several molecules that originate from the immune system have acquired novel functions in the mammary gland. For example, XDH is a highly conserved enzyme with a role in innate immunity (Vorbach et al., 2003), and it is also an essential component of the milk fat globule, and is required for lactation (Vorbach et al., 2002). The regulation of *Xdh* by C/EBPβ in immune cells may have been critical for response to infections. However, as XDH acquired new roles in mammary gland development, so too have it's transcriptional regulators, including C/EBPβ. Lysozyme is another example of a factor with antimicrobial properties that is also found in milk (Vorbach et al., 2006). Furthermore, the α-lactalbumin gene, which encodes a nutritional component of milk, evolved from the lysozyme gene following a gene duplication event that occurred at the start of mammalian evolution approximately 200 million years ago (Prager and Wilson, 1988). Further support for this hypothesis that the mammary gland evolved from mucosal epithelium is provided by the observation that inflammatory signalling pathways are key regulators of mammary gland development (Vorbach et al., 2006). For example, the RANKL/NF-κβ pathway is conserved from invertebrates to mammals, and functions primarily to control the expression of antimicrobial molecules, cytokines and co-stimulatory molecules (Vorbach et al., 2006). In the mammary gland, this signalling axis has developed a novel function, whereby it is required for the development of the secretory epithelium (Srivastava et al., 2003), and for restructuring the mammary gland following lactation (Ormandy et al., 2003; Baxter et al., 2006).

Together, these analyses demonstrate that C/EBPβ regulates specific cellular functions during alveolar development, including signal transduction through regulating the expression of genes involved in extracellular matrix functions, and chromosomal packaging for cell division by regulating genes involved in microtubule assembly, chromosome condensation and centromere formation. As well, C/EBPβ regulates genes that have traditionally been associated with immune

functions. However, as exemplified by XDH, novel roles for C/EBPβ-regulated genes in alveolar development are being identified.

4.5.5 C/EBPβ promotes alveolar development in part through regulation of Elf5

Several novel C/EBPβ targets were identified in luminal progenitor cells that have key roles in alveolar development, and this prompted a deeper investigation of their relationship with C/EBPβ. One of these target genes is *Elf5*. ELF5 is essential for alveolar differentiation during late pregnancy (Oakes et al., 2008), and recently, it was also shown to play a critical role in specifying the luminal cell fate and regulating progenitor/stem cell function (Chakrabarti et al., 2012). Conditional deletion of *Elf5* from the mammary gland leads to a block in luminal progenitor cell specification, as shown by the loss of $CD61^{pos}$ luminal cells, along with a concomitant increase in myoepithelial-like cells that co-express both basal (K14) and luminal (K8) keratins. As well, the expression of several Notch signalling components is significantly altered (e.g., *Notch4*, *Hes1*, *Hey1*, *Cbf1*) upon deletion of *Elf5* (Chakrabarti et al., 2012). Over-activation of Notch signalling was previously shown to lead to differentiation of MaSCs towards the ductal luminal fate, whereas inhibition of the canonical Notch effector *Cbf1* led to expansion of MaSCs and increased stem cell activity (Bouras et al., 2008). In *Cebpb*-/- mice, *Elf5* expression is also significantly decreased (Fig. 3.8), and the resultant changes in Notch signalling may be the underlying mechanism for the reprogramming we observed in alveolar progenitor cells (Chapter 3). Further to this, the ChIP-Seq data also shows that some Notch signalling components are direct targets of C/EBPβ (e.g., *Hey1*, *Jag2I*, data not shown), as well as some known ELF5 target genes (e.g., *Ccnd2*, *Muc4*, *Igfb4*, Figs. 4.15 & 4.16). These observations raise the possibility that C/EBPβ mediates alveolar lineage specification in part through the direct regulation of *Elf5* and (some of) its target genes, and in part directly through the regulation of Notch signalling components. A preliminary experiment to lend support to this will be the assessment of the changes in the expression of *Hey1*, *Jag2 Ccnd2*, *Muc4* and *Igfb4* in *Cebpb*-/- mammary tissue.

In addition to C/EBPβ and ELF5, STAT5 is also likely to play a central role in luminal lineage specification. Activation of STAT5 is essential for the polarization of alveolar cells as well as their up-regulation of milk protein gene expression, and their survival (Cui et al., 2004; Anderson et al., 2007). More recently, it was demonstrated that STAT5 is required for the generation and maintenance of luminal progenitor cells (Yamaji et al., 2009). The expression of *Stat5* in *Cebpb*-/- mice is not significantly affected (data not shown), and the ChIP-Seq data demonstrated that C/EBPβ is not bound to *Stat5a*, the isoform most pertinent in mammary gland development (data not shown). However, ELF5 is involved in the direct regulation of *Stat5* expression (Choi et al., 2009), and may provide a point of convergence for these three essential regulators of luminal specification and alveolar development.

In summary, this chapter described the genome-wide binding characteristics of C/EBPβ in $Sca1^{pos}$ cells, $Sca1^{neg}$ cells and P16.5 alveolar cells. These experiments revealed that genome-wide C/EBPβ occupancy is correlated with alveolar cells fate, and that C/EBPβ target genes perform distinct cellular functions in alveolar cells, but not in ductal cells. Furthermore, direct regulation of *Elf5* by C/EBPβ may be one mechanism through which C/EBPβ exerts its alveolar cell fate programming.

5 Concluding remarks and future perspectives

In this thesis, I have begun to investigate the molecular mechanisms through which C/EBPβ regulates postnatal mammary gland development. In Chapter 2, I have described the development and optimization of a flow cytometric staining strategy that enabled a more precise isolation of Sca1^{pos} and Sca1^{neg} luminal progenitor cells. This strategy was then employed in the isolation and characterization of these progenitor cells, and analysis of their global gene expression pattern demonstrated that Sca1^{pos} luminal progenitor cells contain ductal luminal progenitor cells (DLPs), and Sca1^{neg} luminal progenitor cells contain alveolar luminal progenitor cells (ALPs). Furthermore, the *in vivo* genetic priming of Sca1^{pos} and Sca1^{neg} luminal progenitor cells was maintained when the cells were cultured under differentiation conditions *in vitro*, and their colony morphology was consistent with their *in vivo* genetic priming, demonstrating the functional importance of lineage priming in progenitor cells. In Chapter 3, I have shown flow cytometric analyses that demonstrate that C/EBPβ is required for the appropriate specification of ductal and alveolar lineages in the luminal cell compartment. Furthermore, through gene expression analysis, I have described that in the absence of C/EBPβ, alveolar lineage priming is lost, and ductal lineage priming is up-regulated in both Sca1^{pos} and Sca1^{neg} luminal cells, demonstrating that C/EBPβ is important for alveolar-lineage specification. Preliminary data also shows that in addition to severe proliferation defects, the changes in *in vivo* lineage priming in *Cebpb*-/- mice also affect the *in vitro* differentiation potential of *Cebpb*-/- Sca1^{pos} and Sca1^{neg} luminal progenitor cells. Lastly, in Chapter 4, I have described the genome-wide binding characteristics of C/EBPβ in Sca1^{pos}, Sca1^{neg} and P16.5 alveolar cells through ChIP-Seq experiments. These experiments revealed that genome-wide C/EBPβ occupancy is correlated with alveolar cells fate, and that C/EBPβ target genes perform distinct cellular functions in alveolar cells throughout their development. Furthermore, I showed that *Elf5*, a master regulator of alveolar development, is directly regulated by

C/EBPβ, and posit that direct regulation of *Elf5* by C/EBPβ may be one mechanism through which C/EBPβ exerts its alveolar cell fate programming.

The work presented in this thesis contributes significantly to our understanding of mammary gland development. The identification of ductal- and alveolar-programmed luminal cell populations will allow for better definition of the lineage commitment process and its regulation by extrinsic and intrinsic factors, including C/EBPβ, PRL, STAT5 and ELF5. In addition, studies of mammary carcinoma cells-of-origin will now be better informed, as specific lineage tracing of luminal progenitor subsets can be achieved based on the genes differentially expressed in $Sca1^{pos}$ and $Sca1^{neg}$ luminal cells. As well, the work in this thesis makes a significant contribution to our understanding of C/EBPβ function during mammary gland development. The identification of direct, *in vivo* target genes in three luminal cell populations has enabled us to begin to investigate the mechanisms through which C/EBPβ regulates luminal lineage commitment and alveolar specification. For example, the identification of *Elf5* as a direct C/EBPβ target gene has already prompted further investigations into the relationship between C/EBPβ and ELF5. ChIP-Seq experiments to identify ELF5 target genes in $Sca1^{pos}$, $Sca1^{neg}$ and P16.5 alveolar cells in the presence and absence of C/EBPβ (*Cebpb*-/- mice) are currently in progress (one preliminary ChIP experiment was shown in Chapter 4) and will be pursued in my next post with Prof. Nerlov. Further to these experiments, preliminary ChIP experiments to optimize immunoprecipitation of STAT5 and GATA3 have also been initiated, and will be continued. Although *Stat5* and *Gata3* are not direct C/EBPβ target genes, their critical role in alveolar development suggests that they may function within the same regulatory network as C/EBPβ and ELF5. In fact, STAT5 was recently shown to regulate *Elf5* expression (Yamaji et al., 2009), supporting this hypothesis. Elucidating the relationship between these transcription factors (and perhaps others), and their target genes, will provide critical insight into the regulatory networks that govern alveolar development. Establishing the transcriptional hierarchy, and understanding the regulatory relationships between each of these transcription factors provides an exciting challenge for the future.

6 Materials and methods

6.1 Animals

6.1.1 Mouse strains

Mice designated as wild type (WT) are strain C57 black 6 CGR obtained from the Roger Land and Scottish Centre for Regenerative Medicine (SCRM) animal units. *Cebpb*-/- mice have been described (Sterneck et al., 1997). Briefly, chimeras were generated from embryonic stem cells in which part of the *Cebpb* promoter and the full *Cebpb* coding region were replaced by a neomycin resistance gene. Adult virgin female mice were analyzed between the ages of 10-14 weeks, whereas pregnant mice were taken on day 16.5 following plug confirmation.

6.1.2 Genotyping of Cebpb mice

Genomic DNA was extracted from tail samples using 700 µl of Tail Lysis Buffer (50 mM Tris pH 8.0, 100 mM ethylenediaminetetraacetic acid [EDTA], 100 mM NaCl, 1 % (weight to volume [w/v]) SDS, 10 units Proteinase K [Sigma]) for 12 to 16 hours at 56 ^{o}C. Following lysis, 200 µl of NaCl was added, briefly mixed, and centrifuged for 10 minutes at maximum speed (≥ 16,000 x g) at room temperature. The supernatant was transferred to a new 1.5 ml Eppendorf tube, 800 µl of isopropanol was added, the sample was briefly mixed, and centrifuged for 10 minutes at maximum speed at 4 ^{o}C. The DNA pellet was washed with 500 µl of 70 % ethanol and centrifuged for 5 minutes at maximum speed at 4 ^{o}C, and then dired for 10 minutes. The DNA sample was resuspended in 100 µl of sterile water.

The genotyping PCR reaction was comprised of 4 µl of genomic DNA, 1 µl of each of the following 10 µM primers: G22 5'acgagactagtgagacgtgctac, G23 5' gcttcgaacccgcggactgcaa and G24 5'catctttaaggtgattactcagggc, 2 µl of 2 mM deoxyribonucleotide triphosphates [dNTPs] (Roche), 4 µl of 5x Colorless GoTaq

Flexi Buffer (Promega), 1.2 µl of 25 mM $MgCl_2$ (Promega), 2.5 units (0.5 µl) of GoTaq DNA polymerase (Promega), 2 µl of dimethyl sulfoxide [DMSO], and 3.3 µl of sterile water. Amplification was performed using a DNA Engine Thermal Cycler (BioRad) as follows: denaturation for 5 minutes at 95 ^{o}C, 5 cycles of: 1 minute at 95 ^{o}C, 1 minute at 50 ^{o}C and 1 minute at 72 ^{o}C, followed by 35 cycles of: 1 minute at 95 ^{o}C, 1 minute at 55 ^{o}C and 1 minute at 72 ^{o}C, and a final extension for 5 minutes at 72 ^{o}C. Amplicons were separated on a 2 % (w/v) agarose gel and visualized on a Syngene G:BOX (Fisher). The wild type allele generated a 771 bp amplicon, and the mutant allele generated a 530 bp amplicon.

6.2 Flow cytometry

6.2.1 Mammary gland dissection and single cell preparation

The four limbs of a female mouse killed by a Schedule 1 method were pinned, and a midline incision was created from the vagina to the jaw. After pulling back the skin to expose the subcutaneous mammary fat pads, all five glands were dissected from each side of the animal and placed in 4 ml of cold Dulbecco's Modified Eagle Medium (DMEM)/F-12 (Gibco). The tissue was transferred onto a glass plate, minced to 1-2 mm fragments using two scalpels, transferred to a 15 ml conical tube, and digested in 4 ml of DMEM/F-12 containing 2 mg/ml Collagenase D (Roche) and 6 units Dispase II (Roche) for 1-2 hours at 37 ^{o}C on a rotating shaker (40 to 50 rpm). The sample was then washed with 10 ml of sterile, cold HEPES buffer, pH 7.5 (20 mM HEPES, 140 mM NaCl, 10 mM KCl, 0.1 % (w/v) D-glucose, 0.5 % (w/v) bovine serum albumin [BSA], 2 mM EDTA, 2 mM $CaCl_2$) and centrifuged for 5 minutes at 300 x g at 4 ^{o}C. The cell pellets were resuspended in 1 ml of GEY solution, pH 7.0 (150 mM NH_4Cl, 10 mM $KHCO_3$, 0.1 mM EDTA), incubated for 3 to 5 minutes at room temperature, washed with 14 ml of cold HEPES buffer, and centrifuged for 5 minutes at 300 x g at 4 ^{o}C. The sample was resuspended in 1 ml of HEPES buffer, filtered using a 40 µm Filcon filter (Becton Dickinson) to obtain a single-cell suspension, and the cells counted with a hemocytometer.

6.2.2 Antibody staining

All incubation steps were carried out at 4 ^{o}C for 20 minutes in the dark. To wash the cells following each incubation, the suspension was diluted between 1:10 to 1:20 with cold HEPES buffer, centrifuged for 5 minutes at 300 x g at 4 ^{o}C, and the supernatant was discarded. Dissociated mammary gland cells were first incubated with lineage antibodies (CD31, CD45, Ter119), then a goat anti-rat secondary antibody, followed by a blocking step with normal rat serum. The cells were then incubated with CD29-biotin, and lastly with a cocktail containing streptavidin (SA), CD24, CD49f, Sca1 and CD49b. Further information about the antibodies is found in Table 6.1. After completion of the staining protocol, the cells were resuspended in cold HEPES buffer containing the viability marker 7-AAD (1 ng/ml, Cambridge Bioscience) and filtered with a 40 μm Filcon filter (Becton Dickinson) prior to flow cytometric analysis or cell sorting.

Table 6.1 Antibodies used for flow cytometry.
The table lists the antibodies used for the detection of cell surface antigens using flow cytometry. The incubation steps in the protocol are indicated on the left, followed by the antibody name and directly conjugated fluorophore. The lineage antibodies (step 1) are unconjugated, and were detected with a goat anti-rat secondary antibody (step 2). The dilution of each antibody is shown, as well as clone name, supplier, stock concentration and isotype control.

Step	Antibody & fluorophore	Clone	Supplier	Stock	Dilution	Isotype
1	CD31 Pure	390	eBioscience	0.5 mg/ml	1:450	Rat IgG2a
	CD45 Pure	30-F11	eBioscience	0.5 mg/ml	1:450	Rat IgG2b
	Ter119 Pure	TER119	eBioscience	0.5 mg/ml	1:450	Rat IgG2b
2	Goat anti-rat IgG $F(ab')_2$ PE-Cy5	-	SCBT	0.4 mg/ml	1:100	-
3	Normal rat serum	-	Sigma	-	-	-
4	CD29 Biotin	Ha2/5	BD	0.5 mg/ml	1:100	Hamster IgM
5	SA AF750-APC	-	Invitrogen	1 mg/ml	1:300	-
	CD24 FITC or PB	M1/69	Biolegend	0.5 mg/ml	1:150	Rat IgG2b
	CD49f PE or FITC	GoH3	Biolegend	0.2 mg/ml	1:150	Rat IgG2a
	Sca1 PE-Cy7	D7	eBioscience	0.2 mg/ml	1:300	Rat IgG2a
	CD49b APC	HMα2	BD	0.2 mg/ml	1:100	Hamster IgG1

6.2.3 Cell sorting and analysis

Single cell suspensions stained as described above were analyzed using an LSRFortessa II cell analyzer (Becton Dickinson), or sorted using either a FACSAria II (Becton Dickinson) or FACSJAZZ (Becton Dickinson). All flow cytometric data analyses were performed using FlowJo version 8.7.1 (Tree Star Inc).

6.3 Mouse mammary colony-forming cell assay

6.3.1 Preparation of fibroblast cell layer

NIH3T3 fibroblast cells (ATCC) were plated on a 48-well plate (Iwaki, 10,000 cells/well) in 400 µl of complete Iscove's Modified Dulbecco's Medium (IMDM) (IMDM [Gibco], 5 % fetal calf serum (FCS) [Gibco], 10 units/ml penicillin and 10 µg/ml streptomycin [P/S, Invitrogen]), and incubated in a tissue culture incubator (37 ^{o}C, 5 % CO_2). After 24 hours, the medium was removed and replaced with complete IMDM containing 4 µg/ml of mitomycin-C (Sigma). The cells were incubated with mitomycin-C for 4 hours at 37 ^{o}C with 5 % CO_2, then washed 3 x with 400 µl of warm phosphate buffered saline (PBS, Gibco) and 1 x with 400 µl of complete IMDM. Alternately, cells were irradiated using standard methods. Finally, the cells were incubated for 24 hours in complete IMDM medium at 37 ^{o}C with 5 % CO_2.

6.3.2 Three-dimensional extracellular matrix assay

Two thousand (2000) sorted mammary gland cells were embedded in 100 µl of chilled Matrigel (Becton Dickinson) and plated on top of the mitomycin-C treated fibroblast cell layer. To promote solidification of the Matrigel, the plate was incubated for 10 minutes at 37 ^{o}C with 5 % CO_2. For assessing alveolar differentiation and ductal branching, the cells were incubated with 400 µl of mouse Epicult-B medium [Stem Cell Technologies], 1x Epicult supplement [Stem Cell Technologies], 1x B27 supplement (Gibco), 10 ng/ml recombinant human epidermal growth factor [rh EGF, Stem Cell Technologies], 10 ng/ml recombinant human basic

fibroblast growth factor [rh bFGF, PeproTech], 2 mM glutamax, 10 μM ROCK inhibitor Y-27632 (Calbiochem), 5 % FCS, 1x gentamycin. Fresh medium was added after 24 hours without ROCK inhibitor and FCS and changed at 3-day intervals. From day 7-15 of culture, the medium was supplemented with prolactin (3 μg/ml, Sigma) and RANKL (200 ng/ml, R&D Systems). Colonies were scored on day 15 of culture. For assessing ductal morphogenesis, the cells were incubated with 400 μl of mouse Epicult-B medium, 1x Epicult supplement, 10 ng/ml rh EGF, 10 ng/ml rh bFGF, 4 μg/ml Heparin [Stem Cell Technologies], 5 % FCS, 10 units/ml P/S. The medium was changed at three days intervals, and the colonies were scored after 21 days in culture.

6.4 Chromatin immunoprecipitation sequencing

6.4.1 Cross-linking, lysis and sonication of mammary cells

Sorted cells (see Section 6.2) were centrifuged for 10 minutes at 300 x g at 4 ^{o}C, and the supernatant was removed. The cell pellets were gently resuspended in 1 ml of DMEM/F-12 containing 1 % formaldehyde (Sigma) and incubated for 10 minutes with gentle agitation at room temperature. The fixation was stopped with 125 mM glycine for 5 minutes with gentle agitation at room temperature. The samples were then centrifuged for 5 minutes at 300 x g at 4 ^{o}C, and washed 2 x with 1 ml of Hank's Balanced Salt Solution (HBSS, Invitrogen) containing cOmplete Mini EDTA-free Protease Inhibitor Cocktail Tablet (PIC, Roche, prepared according to the manufacturer's instructions).

The cell pellets were resuspended in ChIP Lysis Buffer (50 mM Tris, pH 8.0, 10 mM EDTA, 1 % SDS, PIC) and incubated for 20 minutes on ice. For pellets containing less than $5x10^5$ cells, 200 μl of ChIP Lysis Buffer was used, and for pellets containing between $5x10^5$ to $2x10^6$ cells, 500 μl of ChIP Lysis Buffer was used.

The samples were sonicated on a high setting for 50 cycles of 30 seconds on and 30 seconds off using a Bioruptor NextGen Sonicator (Diagenode) to obtain DNA

fragments between 200-400 base pairs in size. The sheared chromatin was diluted with RIPA Buffer (10 mM Tris pH 7.5, 1 mM EDTA, 1 % Triton X-100, 0.1 % SDS, 0.1 % sodium deoxycholate, 100 mM NaCl, PIC) so that 400 µl contained chromatin from $1x10^5$ cells. Aliquots of chromatin from $1x10^5$ cells/tube were prepared in 1.5 ml low retention tubes (Axygen), and either frozen and stored at -80 ^oC, or used immediately for preclearing and antibody binding.

6.4.2 Preclearing, antibody binding and antibody capture

Magnetic Dynabeads Protein A (Invitrogen) was used for the removal of non-specific binding to protein A beads (preclearing). To prepare the beads for chromatin preclearing, 20 µl of slurry per sample was washed three times with cold RIPA Buffer, resuspended in the original volume with RIPA Buffer, added to each aliquot of sonicated chromatin, and incubated for one hour with moderate rotation (20 rotations per minute [rpm] on a Stuart SB3 rotator) at 4 ^oC. The beads were then captured with a DynaMag-2 magnetic tube holder (Invitrogen), and the cleared chromatin (supernatant) was transferred to a new 1.5 ml low-retention tube.

Immunoprecipitation was performed with 0.5 µg of anti-C/EBPβ (C19, Santa Cruz Biotechnology), 0.5 µg of anti-Elf5 (H62, Santa Cruz Biotechnology; AP7394C, Abgent), or 0.5 µg of rabbit IgG (Santa Cruz Biotechnology) for 12-14 hours (overnight) at 20 rpm at 4 ^oC. Concomitantly, another aliquot of Dynabeads Protein A beads (20 µl/sample) was blocked in 1 ml of RIPA Buffer containing 10 µg of BSA for 12-14 hours with 20 rpm rotation at 4 ^oC. The following day, the supernatant was removed from the beads, and they were resuspended in their original volume with RIPA Buffer. The blocked beads (20 µl/sample) were then incubated with the samples for 2 hours with 20 rpm rotation at 4 ^oC to capture the immunocomplexes.

Subsequently, the protein A-captured immunocomplexes were washed 3 x with cold RIPA Buffer, 4 x with cold RIPA500 Buffer (RIPA buffer containing 500 mM NaCl), 1 x with cold Lithium Chloride Buffer (250 mM LiCl, 10 mM Tris, pH 8.0,

1 mM EDTA, 0.5 % nonyl phenoxypolyethoxylethanol [NP40], 0.5 % SDS), and 2 x with cold TE Buffer (10 mM Tris, pH 8.0, 10 mM EDTA). Each wash was incubated for 4 minutes at 20 rpm at room temperature. Finally, RNA contaminants were removed by incubation with 50 µg/ml RNAse A (Sigma) for 30 min at 37 ^{o}C.

6.4.3 Reverse cross-linking and DNA purification

Following the washes and RNAse-A treatment, the protein A-bound immunocomplexes were resuspended in 200 µl of ChIP Elution Buffer (20 mM Tris, pH 7.5, 5 mM EDTA, 50 mM NaCl, 1 % SDS, 50 µg/ml Proteinase K [Roche]) and incubated for 2 hours at 68 ^{o}C and 1300 rpm on an Eppendorf Thermomixer. The supernatant containing immunoprecipitated genomic DNA was transferred to a new low-retention tube (Axygen), and the beads were incubated with a further 100 µl of ChIP Elution Buffer for 10 minutes to remove any remaining DNA. The supernatant was pooled with the previous aliquot for each sample, and the beads were discarded.

One additional chromatin sample (referred to as 'input'), which did not undergo immunoprecipitation, was reverse cross-linked by supplementing the RIPA Buffer with Proteinase K (to a final concentration of 50 µg/ml) and SDS (to a final concentration of 1 %) and incubating for 2 hours at 68 ^{o}C and 1300 rpm on an Eppendorf Thermomixer along with the ChIP samples. The volumes of the reagents used in the subsequent phenol-chloroform extraction were also adjusted for the input sample to reflect its larger volume of approximately 440 µl.

The reverse cross-linked genomic DNA was purified by phenol-chloroform extraction. To begin, samples were mixed with one volume (300 µl) of phenol-chloroform-isoamyl-alcohol (Invitrogen), transferred to 2 ml phase-lock tubes (VWR), and centrifuged for 5 minutes at maximum speed at room temperature. The aqueous phase, separated from the organic phase by the phase-lock gel in the tubes, was then mixed with one volume (300 µl) of chloroform (Invitrogen) and centrifuged for 5 minutes at maximum speed at room temperature. The remaining aqueous phase was transferred to a fresh 1.5 ml low-retention tube, and the DNA was precipitated

overnight at -20 ^{o}C with 2.5 volumes (750 µl) of ethanol in the presence of 50 µg glycogen (Ambion) and 30 µl of 3 M sodium acetate, pH 5.2. The following day, the samples were centrifuged for 20 minutes at maximum speed at 4 ^{o}C, washed with 500 µl of 70 % ethanol, centrifuged for 5 minutes at maximum speed at 4 ^{o}C, air dried for 10-15 minutes, and resuspended in 20 µl of sterile water.

6.4.4 ChIP validation using qPCR

Published and putative C/EBPβ and ELF5 target genes were used for initial validation of chromatin immunoprecipitation. *In silico* analysis of putative C/EBPβ and ELF5 binding was performed using TRANSFAC/ExPlain 3.0 (BioBase). qPCR assays were designed against published/putative bound and unbound regions in each gene using the Universal ProbeLibrary (UPL) Assay Design Centre (Roche) (Table 6.2).

Each qPCR reaction contained 2 µl of gDNA, 5 µl of 2 x Probes Master (Roche), 2.6 µl of Probes Water (Roche), 0.2 µl of 10 µM sense and antisense primer mix, and 0.2 µl of 10 µM mouse UPL probe (assay specific). Amplification was performed on a LightCycler 480 (Roche) with the following conditions: denaturation for 10 minutes at 95 ^{o}C, 45 cycles of: 30 seconds at 95 ^{o}C, 10 seconds at 60 ^{o}C and 1 second at 72 ^{o}C (acquisition).

Table 6.2 qPCR assays for validation of C/EBPβ and ELF5 ChIP.
The qPCR primers listed in the table were designed using the Roche UPL assay design centre. The table lists the primer ID (which includes the gene name), the primer sequences (5' to 3') and the accompanying UPL probe. As well, the genomic position of each qPCR assay relative to the transcription start site (TSS) is listed, along with the reference that identified the target gene. Whether the transcription factor of interest is expected to be bound based on published or *in silico* analyses is also shown.

Gene/ Primer ID	Primer sequences (5' to 3')	UPL Probe	Position (TSS=1)	Expected binding	Reference
Xdh-1	F: acagagtaggatctgtggattgg R: ggcaagagtcacaggtttgc	93	-149/-84	C/EBPβ +	Seymour et al., 2006
Xdh-2	F: tgcaaacctgtgactcttgc R: tgaaacaaagttagccaatgagaa	1	-108/-31	C/EBPβ +	
Xdh-3	F: cagagctagcaggcaatctga R: tcttgaagaaggaaatgaatctga	99	+4734/ +4804	C/EBPβ -	
Xdh-4	F: cgagaaatggcacactaaggt R: ggcaaataaaaatgacccaaa	79	+6595/ +6660	C/EBPβ -	
Id2-1	F: ttcgcctggtatgatggac R: gcgtcttttatgtgcactcg	31	-72/-18	C/EBPβ +	Karaya et al., 2005
Id2-2	F: ctcgcgcttcattctgaac R: tgaaggctttcatgctgct	69	-8/+95	C/EBPβ +	
Id2-3	F: cttactatgaaggtgtcaggacca R: ggttaccacctcagtgtaccg	5	+1011/ +1084	C/EBPβ -	
Id2-4	F: gaaacgcaacacacctagca R: cgtttaaaatctagtcccaatgc	21	+2438/ +2498	C/EBPβ -	
Gata3-1	F: ggcgccgtcttgatagttt R: tgtctcgctctctcagtctctct	55	+25/ +125	C/EBPβ +	Unpublished
Gata3-2	F: atagagagctacgcaatctgacc R: acctgagtagcaaggagcgta	63	+125/ +198	C/EBPβ +	
Ccnd2-1	F: tttgtgggttcccctatcc R: tacttctccgaggccctctc	42	+2727/ +2794	ELF5+	Escamilla-Hernandez et al., 2010
Ccnd2-2	F: gcaggcagaagacaggaact R: tctggggtcaggtttctcc	88	+22874/ +22936	ELF5-	
Igfbp4-1	F: tggcacacgaaggaagaga R: gaaggggtctgagaaagctctac	7	+994/ +1069	ELF5+	Escamilla-Hernandez et al., 2010
Igfbp4-2	F: agctttctcagacccccttca R: agcctccccagtctccat	4	+1051/ +1114	ELF5+	
Igfbp4-3	F: tggggaggcttaccaaagta R: cggccaaaggagtacatctg	68	+1152/ +1212	ELF5+	
Stat5a-1	F: ccaagcagaactcagataggtg R: tcccagcaagtgtttgtgtc	108	-2907/ -2813	ELF5-	Choi et al., 2009
Stat5a-2	F: aaatgtgacacttccgggtaa R: ggtgacagaggtgtctgtgg	109	-1451/ -1390	ELF5-	
Stat5a-3	F: gatggagggcatgaccag R: ggtttcctaagttctctgagctacat	13	-769/ -707	ELF5-	
Stat5a-4	F: tactgcccatgccaacttc R: aagtctgccatccccttg	82	-8/+67	ELF5+	
Stat5a-5	F: tgcaggcagatgtgaagttag R: cacctgaagctggtcctacc	4	+173/ +249	ELF5+	

Table 6.2 Continued

Gene/ Primer ID	Primer sequences (5' to 3')	UPL Probe	Position (TSS=1)	Expected binding	Reference
Elf5-1	F: gccctgactgtctggctaaa R: cagagcttgtcagcacttca	-	-12541/ -12387	C/EBPβ -	Unpublished
Elf5-2	F: tattgccctccctctctcag R:gagcacaggagaaatgtgtca	-	-12463/ -12364	C/EBPβ -	
Elf5-3	F: ctcacttttatgaataacgta R: aatttaaataatcactgtggag	-	-6995/ -6846	C/EBPβ -	
Elf5-4	F: tcacttttatgaataacgtat R: gcctatcaatttaaataatcac	-	-6994/ -6839	C/EBPβ -	
Elf5-5	F: tggcactgctcttctttcct R: ccaagtcccgaacctctgta	-	-370/ -301	C/EBPβ +	
Elf5-6	F: caaaggctgcaatgaacaga R: tggagcagttctcgttgag	-	-290/ -220	C/EBPβ +	
Elf5-7	F: acccaggaaatgagcagaga R: agcttgctcaacggagactg	-	-244/ -220	C/EBPβ +	

6.4.5 Preparation of DNA sequencing libraries

Immunoprecipitated DNA from 5-6 sample tubes (equivalent to chromatin from 5-6x10^5 cells) was pooled for preparation of each DNA sequencing library. The starting material and sequencing libraries were quantified using the Quant-iT dsDNA HS Assay (Invitrogen) on either a Nanodrop 3000 (Fisher) or Qubit (Invitrogen) fluorospectrometer.

DNA sequencing libraries were prepared using the Illumina ChIP-Seq Sample Preparation Kit (Illumina) according to the manufacturer's instructions with the exception of the following three changes: 1) Two ng of DNA were used for library preparation instead of 10 ng. 2) Each Qiagen column purification step was modified at the elution stage to maximize DNA recovery. The modification includes heating of Buffer EB to 65 ^{o}C, 3 minutes incubation of Buffer EB on the column prior to centrifugation for 1 minute at maximum speed at room temperature. Reapplication of the eluate onto the column for another 3 minutes incubation, followed by a final spin for 1 minute at maximum speed at room temperature. 3) During the library size selection step, the excised gel fragments were incubated in 6 volumes of Buffer QG for 30 minutes at room temperature instead of 3 volumes of Buffer QG for 10 minutes at 50 ^{o}C.

6.4.6 Solexa sequencing and bioinformatics analysis

DNA sequencing libraries were submitted to the University of Edinburgh GenePool (Edinburgh, Scotland) for Solexa sequencing using the GAIIx Sequencer (Illumina). Samples were verified using the High Sensitivity DNA Chip (Agilent) on a Bioanalyzer 2100 (Agilent) prior to Solexa sequencing. Dr. Simon Tomlinson curated the ChIP-Seq data using Geneprof. Fifty base pair single-end reads were trimmed to remove adapter sequences and aligned to the mouse genome (Ensembl 58 Mouse genome NCBIM37 Assembly) using Bowtie. For quality control, the mean quality score was calculated for each data set by summing up all of the quality scores for each read and dividing the sum by the length of the read. A mean quality score of

≥8 passed the quality control filter. Next, peak calling was performed using MACS (Zhang et al., 2008), and peaks were annotated to genes using the criteria of -20 kb to +1 kb of TSS (default parameters on Geneprof) (Halbritter et al., 2012). Functional annotation clustering analyses were performed using DAVID bioinformatics resource (http://david.abcc.ncifcrf.gov) using the functional annotation tool with default parameters.

6.5 Microarray and qRT-PCR analysis

6.5.1 RNA extraction from sorted mammary epithelial cells

Mammary gland cells (10,000 per array) prepared as described in section 6.2 were sorted directly into 800 µl of Trizol Reagent (Invitrogen). Total RNA was extracted according to the manufacturer's instructions, with the exception that during the precipitation step, 10 µg of linear polyacrylamide (Ambion) were added. RNA was quantified using the Quant-iT RNA Assay (Invitrogen) on a Qubit (Invitrogen) fluorospectrometer. Three biological replicates for $Sca1^{pos}$ and $Sca1^{neg}$ cells, and four biological replicates for P16.5 were used for microarray analysis.

6.5.2 Gene expression using microarray and bioinformatics analysis

RNA samples were submitted to the RH Microarray Center (Copenhagen, Denmark) for analysis on a GeneChip Mouse Exon 1.0 ST Array (Affymetrix). Dr. Simon Tomlinson curated the microarray data using GeneProf (Halbritter et al., 2012). Intensity data were background corrected and normalized at the gene level (core probe sets only) using Robust Multiarray Analysis (RMA), and transcript clusters were annotated using Expression Console. The resulting signals were filtered to exclude Transcript Cluster IDs with minimal variation across samples. Filtered data were tested for differential expression using Limma in the BioConductor framework (Gentleman et al., 2004). Transcript cluster IDs with a false detection rate of $p \leq 0.01$ were considered differentially expressed. Functional annotation clustering analyses were performed using DAVID bioinformatics resource

(http://david.abcc.ncifcrf.gov) using the functional annotation tool with default parameters.

6.5.3 Gene expression using qRT-PCR

Following RNA extraction as described in section 6.5.1, cDNA was synthesized using the Superscript VILO cDNA Synthesis Kit (Invitrogen) according to the manufacturer's instructions. Gene expression was analyzed either using the UPL system on a LightCycler 480 (Roche) (described in section 6.4.4 with the exception that cDNA was used instead of gDNA), or a BioMark 48.48 Dynamic Array (Fluidigm) on a BioMark Genetic Analyzer (Fluidigm) using the manufacturer's instructions.

6.6 Luciferase assays

H293T cells were plated on a 12-well plate (Iwaki) and cultured in DMEM (Gibco) containing 10 % FCS for 24 hours at 37 ^{o}C, 5 % CO_2. After 24 hours, the cells were transfected with pGL3-Elf5-promoter (50 ng), pcDNA3-Cebpb (10-100 ng) or empty vector (pcDNA3, 10-100 ng), and phRL-CMV (500 pg) (Promega) as a transfection control, using the calcium phosphate method. For this, the DNA plasmids were combined with 0.25 M $CaCl_2$, mixed in a 1:1 volume to volume (v/v) ratio with 2x HBS Buffer (50 mM HEPES, 1.5 mM Na_2HPO_4, 280 mM NaCl), and added drop-wise to each well. Transfected cells were harvested after 24 hours, and assayed using the Dual Luciferase Reporter (DLR) Assay (Promega). The cells in each well were lysed in 200 µl of Passive Lysis Buffer (Promega), and 10 ml of lysate was assayed for Renilla and firefly luciferase activity using a GloMax 96 microplate luminometer (Promega) according to the manufacturer's instructions.

7 References

Abell, K., Bilancio, A., Clarkson, R.W.E., Tiffen, P.G., Altaparmakov, A.I., Burdon, T.G., Asano, T., Vanhaesebroeck, B., and Watson, C.J. (2005). Stat3-induced apoptosis requires a molecular switch in PI(3)K subunit composition. Nat Cell Biol *7*, 392–398.

Adams, J., and Watt, F. (1993). Regulation of development and differentiation by the extracellular matrix. Development *117*, 1183-1198.

Akira, S., Isshiki, H., Sugita, T., Tanabe, O., Kinoshita, S., Nishio, Y., Nakajima, T., Hirano, T., and Kishimoto, T. (1990). A nuclear factor for IL-6 expression (NF-IL6) is a member of a C/EBP family. EMBO J. *9*, 1897–1906.

Al-Hajj, M., Wicha, M.S., Benito-Hernandez, A., Morrison, S.J., and Clarke, M.F. (2003). Prospective identification of tumorigenic breast cancer cells. Proc. Natl. Acad. Sci. U.S.A. *100*, 3983–3988.

Alvi, A.J., Clayton, H., Joshi, C., Enver, T., Ashworth, A., Vivanco, M.D.M., Dale, T.C., and Smalley, M.J. (2003). Functional and molecular characterisation of mammary side population cells. Breast Cancer Res *5*, R1–R8.

Anderson, S.M., Rudolph, M.C., McManaman, J.L., and Neville, M.C. (2007). Key stages in mammary gland development. Secretory activation in the mammary gland: it's not just about milk protein synthesis! Breast Cancer Research *9*, 204-218.

Andl, T., Reddy, S.T., Gaddapara, T., and Millar, S.E. (2002). WNT signals are required for the initiation of hair follicle development. Developmental Cell *2*, 643–653.

Andres, A.C., and Strange, R. (1999). Apoptosis in the estrous and menstrual cycles. J Mammary Gland Biol Neoplasia *4*, 221–228.

Ankrapp, D.P., Bennett, J.M., and Haslam, S.Z. (1998). Role of epidermal growth factor in the acquisition of ovarian steroid hormone responsiveness in the normal mouse mammary gland. J. Cell. Physiol. *174*, 251–260.

Antal, M.C., Krust, A., Chambon, P., and Mark, M. (2008). Sterility and absence of histopathological defects in nonreproductive organs of a mouse ERbeta-null mutant. Proc. Natl. Acad. Sci. U.S.a. *105*, 2433–2438.

Asch, H.L., and Asch, B.B. (1985). Expression of keratins and other cytoskeletal proteins in mouse mammary epithelium during the normal developmental cycle and primary culture. Developmental Biology *107*, 470–482.

Asselin-Labat, M.-L., Sutherland, K.D., Barker, H., Thomas, R., Shackleton, M., Forrest, N.C., Hartley, L., Robb, L., Grosveld, F.G., van der Wees, J., et al. (2006). Gata-3 is an essential regulator of mammary-gland morphogenesis and luminal-cell differentiation. Nat Cell Biol *9*, 201–209.

Attema, J.L., Papathanasiou, P., Forsberg, E.C., Xu, J., Smale, S.T., and Weissman, I.L. (2007). Epigenetic characterization of hematopoietic stem cell differentiation using miniChIP and bisulfite sequencing analysis. Proc. Natl. Acad. Sci. U.S.A. *104*, 12371–12376.

Aupperlee, M.D. (2005). Progesterone Receptor Isoforms A and B: Temporal and Spatial Differences in Expression during Murine Mammary Gland Development. Endocrinology *146*, 3577–3588.

Baldwin, B.R., Timchenko, N.A., and Zahnow, C.A. (2004). Epidermal growth factor receptor stimulation activates the RNA binding protein CUG-BP1 and increases expression of C/EBPbeta-LIP in mammary epithelial cells. Molecular and Cellular Biology *24*, 3682–3691.

Bamshad, M., Le, T., Watkins, W.S., Dixon, M.E., Kramer, B.E., Roeder, A.D., Carey, J.C., Root, S., Schinzel, A., Van Maldergem, L., et al. (1999). The Spectrum of Mutations in TBX3: Genotype/Phenotype Relationship in Ulnar-Mammary Syndrome. The American Journal of Human Genetics *64*, 1550–1562.

Barcellos-Hoff, M.H., Aggeler, J., Ram, T.G., and Bissell, M.J. (1989). Functional differentiation and alveolar morphogenesis of primary mammary cultures on reconstituted basement membrane. Development *105*, 223–235.

Baxter, F.O., Came, P.J., Abell, K., Kedjouar, B., Huth, M., Rajewsky, K., Pasparakis, M., and Watson, C.J. (2006). IKKbeta/2 induces TWEAK and apoptosis in mammary epithelial cells. Development *133*, 3485–3494.

Benavides, A., Pastor, D., Santos, P., Tranque, P., and Calvo, S. (2005). CHOP plays a pivotal role in the astrocyte death induced by oxygen and glucose deprivation. Glia *52*, 261–275.

Bernardo, G.M., Lozada, K.L., Miedler, J.D., Harburg, G., Hewitt, S.C., Mosley, J.D., Godwin, A.K., Korach, K.S., Visvader, J.E., Kaestner, K.H., et al. (2010). FOXA1 is an essential determinant of ERalpha expression and mammary ductal morphogenesis. Development *137*, 2045–2054.

Berrier, A., Siu, G., and Calame, K. (1998). Transcription of a minimal promoter from the NF-IL6 gene is regulated by CREB/ATF and SP1 proteins in U937 promonocytic cells. J. Immunol. *161*, 2267–2275.

Birsoy, K., Chen, Z., and Friedman, J. (2008). Transcriptional regulation of adipogenesis by KLF4. Cell Metab. *7*, 339–347.

Bittner, J.J. (1936). Some possible effects of nursing on the mammary gland tumor incidence in mice. Science *84*, 162.

Blanpain, C., and Fuchs, E. (2009). Epidermal homeostasis: a balancing act of stem cells in the skin. Nat Rev Mol Cell Biol *10*, 207–217.

Blanpain, C., Lowry, W.E., Geoghegan, A., Polak, L., and Fuchs, E. (2004). Self-renewal, multipotency, and the existence of two cell populations within an epithelial stem cell niche. Cell *118*, 635–648.

Bloushtain-Qimron, N., Yao, J., Snyder, E.L., Shipitsin, M., Campbell, L.L., Mani, S.A., Hu, M., Chen, H., Ustyansky, V., Antosiewicz, J.E., et al. (2008). Cell type-specific DNA methylation patterns in the human breast. Proc. Natl. Acad. Sci. U.S.A. *105*, 14076–14081.

Bocchinfuso, W.P., Lindzey, J.K., Hewitt, S.C., Clark, J.A., Myers, P.H., Cooper, R., and Korach, K.S. (2000). Induction of mammary gland development in estrogen receptor-alpha knockout mice. Endocrinology *141*, 2982–2994.

Bole-Feysot, C., Goffin, V., Edery, M., Binart, N., and Kelly, P.A. (1998). Prolactin (PRL) and its receptor: actions, signal transduction pathways and phenotypes observed in PRL receptor knockout mice. Endocr. Rev. *19*, 225–268.

Bonnette, S.G., and Hadsell, D.L. (2001). Targeted disruption of the IGF-I receptor gene decreases cellular proliferation in mammary terminal end buds. Endocrinology *142*, 4937–4945.

Boudreau, N., Sympson, C.J., Werb, Z., and Bissell, M.J. (1995). Suppression of ICE and apoptosis in mammary epithelial cells by extracellular matrix. Science *267*, 891–893.

Bouras, T., Pal, B., Vaillant, F., Harburg, G., Asselin-Labat, M.L., Oakes, S.R., Lindeman, G.J., and Visvader, J.E. (2008). Notch signaling regulates mammary stem cell function and luminal cell-fate commitment. Cell Stem Cell *3*, 429–441.

Boutin, J.M., Jolicoeur, C., Okamura, H., Gagnon, J., Edery, M., Shirota, M., Banville, D., Dusanter-Fourt, I., Djiane, J., and Kelly, P.A. (1988). Cloning and expression of the rat prolactin receptor, a member of the growth hormone/prolactin receptor gene family. Cell *53*, 69–77.

Boxer, R.B., Stairs, D.B., Dugan, K.D., Notarfrancesco, K.L., Portocarrero, C.P., Keister, B.A., Belka, G.K., Cho, H., Rathmell, J.C., Thompson, C.B., et al. (2006). Isoform-specific requirement for Akt1 in the developmental regulation of cellular metabolism during lactation. Cell Metab. *4*, 475–490.

Bridges, J.P., Lin, S., Ikegami, M., and Shannon, J.M. (2012). Conditional hypoxia inducible factor-1α induction in embryonic pulmonary epithelium impairs maturation and augments lymphangiogenesis. Developmental Biology *362*, 24–41.

Brisken, C. (2002). Hormonal control of alveolar development and its implications for breast carcinogenesis. J Mammary Gland Biol Neoplasia *7*, 39–48.

Brisken, C., and O'Malley, B. (2010). Hormone action in the mammary gland. Cold Spring Harbor Perspectives in Biology *2*, a003178–a003178.

Brisken, C., Heineman, A., Chavarria, T., Elenbaas, B., Tan, J., Dey, S.K., McMahon, J.A., McMahon, A.P., and Weinberg, R.A. (2000). Essential function of Wnt-4 in mammary gland development downstream of progesterone signaling. Genes & Development *14*, 650–654.

Brisken, C., Kaur, S., Chavarria, T.E., Binart, N., Sutherland, R.L., Weinberg, R.A., Kelly, P.A., and Ormandy, C.J. (1999). Prolactin Controls Mammary Gland Development via Direct and Indirect Mechanisms. Developmental Biology *210*, 96–106.

Brisken, C., Park, S., Vass, T., Lydon, J.P., O'Malley, B.W., and Weinberg, R.A. (1998). A paracrine role for the epithelial progesterone receptor in mammary gland development. Developmental Biology *210*, 96-106.

Bry, C., Maass, K., Miyoshi, K., Willecke, K., Ott, T., Robinson, G.W., and Hennighausen, L. (2004). Loss of connexin 26 in mammary epithelium during early but not during late pregnancy results in unscheduled apoptosis and impaired development. Developmental Biology *267*, 418–429.

Buchsbaum, R., Telliez, J.B., Goonesekera, S., and Feig, L.A. (1996). The N-terminal pleckstrin, coiled-coil, and IQ domains of the exchange factor Ras-GRF act cooperatively to facilitate activation by calcium. Molecular and Cellular Biology *16*, 4888–4896.

Bundy, L.M., and Sealy, L. (2003). CCAAT/enhancer binding protein beta (C/EBPβ)-2 transforms normal mammary epithelial cells and induces epithelial to mesenchymal transition in culture. Oncogene *22*, 869–883.

Cao, Y., Bonizzi, G., Seagroves, T.N., Greten, F.R., Johnson, R., Schmidt, E.V., and Karin, M. (2001). IKKalpha provides an essential link between RANK signaling and cyclin D1 expression during mammary gland development. Cell *107*, 763–775.

Cao, Z., Umek, R.M., and McKnight, S.L. (1991). Regulated expression of three C/EBP isoforms during adipose conversion of 3T3-L1 cells. Genes & Development *5*, 1538–1552.

Chakrabarti, R., Wei, Y., Romano, R.-A., Decoste, C., Kang, Y., and Sinha, S. (2012). Elf5 regulates mammary gland stem/progenitor cell fate by influencing notch signaling. Stem Cells *30*, 1496–1508.

Chang, C.J., Chen, T.T., Lei, H.Y., Chen, D.S., and Lee, S.C. (1990). Molecular cloning of a transcription factor, AGP/EBP, that belongs to members of the C/EBP family. Molecular and Cellular Biology *10*, 6642–6653.

Chang, C.J., Shen, B.J., and Lee, S.C. (1995). Autoregulated induction of the acute-phase response transcription factor gene, agp/ebp. DNA Cell Biol. *14*, 529–537.

Chapman, R.S., Lourenco, P.C., Tonner, E., Flint, D.J., Selbert, S., Takeda, K., Akira, S., Clarke, A.R., and Watson, C.J. (1999). Suppression of epithelial apoptosis and delayed mammary gland involution in mice with a conditional knockout of Stat3. Genes & Development *13*, 2604–2616.

Charafe-Jauffret, E., Ginestier, C., Monville, F., Finetti, P., Adélaïde, J., Cervera, N., Fekairi, S., Xerri, L., Jacquemier, J., Birnbaum, D., et al. (2005). Gene expression profiling of breast cell lines identifies potential new basal markers. Oncogene *25*, 2273–2284.

Cho, K.-W., Kim, J.-Y., Song, S.-J., Farrell, E., Eblaghie, M.C., Kim, H.-J., Tickle, C., and Jung, H.-S. (2006). Molecular interactions between Tbx3 and Bmp4 and a model for dorsoventral positioning of mammary gland development. Proc. Natl. Acad. Sci. U.S.A. *103*, 16788–16793.

Choi, Y.S., Chakrabarti, R., Escamilla-Hernandez, R., and Sinha, S. (2009). Elf5 conditional knockout mice reveal its role as a master regulator in mammary alveolar development: Failure of Stat5 activation and functional differentiation in the absence of Elf5. Developmental Biology *329*, 227–241.

Christian, M., Pohnke, Y., Kempf, R., Gellersen, B., and Brosens, J.J. (2002). Functional association of PR and CCAAT/enhancer-binding protein beta isoforms: promoter-dependent cooperation between PR-B and liver-enriched inhibitory protein, or liver-enriched activatory protein and PR-A in human endometrial stromal cells. Molecular Endocrinology *16*, 141–154.

Chu, E.Y., Hens, J., Andl, T., Kairo, A., Yamaguchi, T.P., Brisken, C., Glick, A., Wysolmerski, J.J., and Millar, S.E. (2004). Canonical WNT signaling promotes mammary placode development and is essential for initiation of mammary gland morphogenesis. Development *131*, 4819–4829.

Ciarloni, L., Mallepell, S., and Brisken, C. (2007). Amphiregulin is an essential mediator of estrogen receptor alpha function in mammary gland development. Proc. Natl. Acad. Sci. U.S.A. *104*, 5455–5460.

Clarkson, R., Wayland, M., and Lee, J. (2004). Gene expression profiling of mammary gland development reveals putative roles for death receptors and immune mediators in post-lactational regression. Breast Cancer Research *6*, R92-R109.

Conneely, O.M., Mulac-Jericevic, B., and Lydon, J.P. (2003). Progesterone-dependent regulation of female reproductive activity by two distinct progesterone receptor isoforms. Steroids *68*, 771–778.

Cooper, C., Henderson, A., Artandi, S., Avitahl, N., and Calame, K. (1995). Ig/EBP (C/EBP gamma) is a transdominant negative inhibitor of C/EBP family transcriptional activators. Nucleic Acids Research *23*, 4371–4377.

Couldrey, C., Moitra, J., Vinson, C., Anver, M., Nagashima, K., and Green, J. (2002). Adipose tissue: a vital in vivo role in mammary gland development but not differentiation. Dev. Dyn. *223*, 459–468.

Creamer, B.A., Sakamoto, K., Schmidt, J.W., Triplett, A.A., Moriggl, R., and Wagner, K.U. (2010). Stat5 Promotes Survival of Mammary Epithelial Cells through Transcriptional Activation of a Distinct Promoter in Akt1. Molecular and Cellular Biology *30*, 2957–2970.

Cremers, N., Deugnier, M.-A., and Sleeman, J. (2010). Loss of CD24 expression promotes ductal branching in the murine mammary gland. Cell. Mol. Life Sci. *67*, 2311–2322.

Croniger, C., Trus, M., Lysek-Stupp, K., Cohen, H., Liu, Y., Darlington, G.J., Poli, V., Hanson, R.W., and Reshef, L. (1997). Role of the isoforms of CCAAT/enhancer-binding protein in the initiation of phosphoenolpyruvate carboxykinase (GTP) gene transcription at birth. Journal of Biological Chemistry *272*, 26306–26312.

Cui, Y., Riedlinger, G., Miyoshi, K., Tang, W., Li, C., Deng, C.X., Robinson, G.W., and Hennighausen, L. (2004). Inactivation of Stat5 in mouse mammary epithelium during pregnancy reveals distinct functions in cell proliferation, survival, and differentiation. Molecular and Cellular Biology *24*, 8037–8047.

Daniel, C., and Silberstein, G. (1987). The Mammary Gland (New York: Plennum Press).

Daniel, C.W., De Ome, K.B., Young, J.T., BLAIR, P.B., and FAULKIN, L.J. (1968). The in vivo life span of normal and preneoplastic mouse mammary glands: a serial transplantation study. Proc. Natl. Acad. Sci. U.S.A. *61*, 53–60.

Darlington, G.J., Ross, S.E., and MacDougald, O.A. (1998). The role of C/EBP genes in adipocyte differentiation. Journal of Biological Chemistry *273*, 30057–30060.

DasGupta, R., and Fuchs, E. (1999). Multiple roles for activated LEF/TCF transcription complexes during hair follicle development and differentiation. Development *126*, 4557–4568.

Davenport, T.G., Jerome-Majewska, L.A., and Papaioannou, V.E. (2003). Mammary gland, limb and yolk sac defects in mice lacking Tbx3, the gene mutated in human ulnar mammary syndrome. Development *130*, 2263–2273.

Davis, J.A., and Linzer, D.I. (1989). Expression of multiple forms of the prolactin receptor in mouse liver. Molecular Endocrinology *3*, 674–680.

Dearth, L., Hutt, J., and Sattler, A. (2001). Expression and function of CCAAT/enhancer binding proteinβ (C/EBPβ) LAP and LIP isoforms in mouse mammary gland, tumors and cultured mammary epithelial cells. Journal of Cellular Biochemistry *82*, 357-370.

Deome, K., Faulkin, L., Bern, H., and Blair, P. (1959). Development of mammary tumors from hyperplastic alveolar nodules transplanted into gland-free mammary fat pads of female C3H mice. Cancer Research *19*, 515–520.

Descombes, P., and Schibler, U. (1991). A liver-enriched transcriptional activator protein, LAP, and a transcriptional inhibitory protein, LIP, are translated from the same mRNA. Cell *67*, 569-579.

Dickson, S.R., and Warburton, M.J. (1992). Enhanced synthesis of gelatinase and stromelysin by myoepithelial cells during involution of the rat mammary gland. J. Histochem. Cytochem. *40*, 697–703.

Dunbar, M., Dann, P., and Robinson, G. (1999). Parathyroid hormone-related protein signaling is necessary for sexual dimorphism during embryonic mammary development. Development *126*, 3485-3493.

Duprez, E., Wagner, K., Koch, H., and Tenen, D.G. (2003). C/EBPbeta: a major PML-RARA-responsive gene in retinoic acid-induced differentiation of APL cells. Embo J. *22*, 5806–5816.

Eaton, E.M., Hanlon, M., Bundy, L., and Sealy, L. (2001). Characterization of C/EBPβ isoforms in normal versus neoplastic mammary epithelial cells. J. Cell. Physiol. *189*, 91–105.

Eblaghie, M.C., Song, S.-J., Kim, J.-Y., Akita, K., Tickle, C., and Jung, H.-S. (2004). Interactions between FGF and Wnt signals and Tbx3 gene expression in mammary gland initiation in mouse embryos. J Anatomy *205*, 1–13.

Escamilla-Hernandez, R., Chakrabarti, R., Romano, R.-A., Smalley, K., Zhu, Q., Lai, W., Halfon, M.S., Buck, M.J., and Sinha, S. (2010). Genome-wide search identifies Ccnd2 as a direct transcriptional target of Elf5 in mouse mammary gland. BMC Mol Biol *11*, 68-84.

Fan, H.-Y., Liu, Z., Johnson, P.F., and Richards, J.S. (2011). CCAAT/enhancer-binding proteins (C/EBP)-α and -β are essential for ovulation, luteinization, and the expression of key target genes. Mol. Endocrinol. *25*, 253–268.

Fan, H.-Y., Liu, Z., Shimada, M., Sterneck, E., Johnson, P.F., Hedrick, S.M., and Richards, J.S. (2009). MAPK3/1 (ERK1/2) in ovarian granulosa cells are essential for female fertility. Science *324*, 938–941.

Fantl, V., Edwards, P.A., Steel, J.H., Vonderhaar, B.K., and Dickson, C. (1999). Impaired mammary gland development in Cyl-1(-/-) mice during pregnancy and lactation is epithelial cell autonomous. Developmental Biology *212*, 1–11.

Fata, J.E., Chaudhary, V., and Khokha, R. (2001a). Cellular turnover in the mammary gland is correlated with systemic levels of progesterone and not 17beta-estradiol during the estrous cycle. Biology of Reproduction *65*, 680–688.

Fata, J.E., Leco, K.J., Voura, E.B., Yu, H.Y., Waterhouse, P., Murphy, G., Moorehead, R.A., and Khokha, R. (2001b). Accelerated apoptosis in the Timp-3-deficient mammary gland. J. Clin. Invest. *108*, 831–841.

Fata, J.E., Werb, Z., and Bissell, M.J. (2004). Regulation of mammary gland branching morphogenesis by the extracellular matrix and its remodeling enzymes. Breast Cancer Res *6*, 1–11.

Fendrick, J.L., Raafat, A.M., and Haslam, S.Z. (1998). Mammary gland growth and development from the postnatal period to postmenopause: ovarian steroid receptor ontogeny and regulation in the mouse. J Mammary Gland Biol Neoplasia *3*, 7–22.

Feng, Y., Manka, D., Wagner, K.-U., and Khan, S.A. (2007). Estrogen receptor-alpha expression in the mammary epithelium is required for ductal and alveolar morphogenesis in mice. Proc. Natl. Acad. Sci. U.S.A. *104*, 14718–14723.

Fernandez-Valdivia, R., Mukherjee, A., Ying, Y., Li, J., Paquet, M., DeMayo, F.J., and Lydon, J.P. (2009). The RANKL signaling axis is sufficient to elicit ductal side-branching and alveologenesis in the mammary gland of the virgin mouse. Developmental Biology *328*, 127–139.

Foley, J., Dann, P., Hong, J., and Cosgrove, J. (2001). Parathyroid hormone-related protein maintains mammary epithelial fate and triggers nipple skin differentiation during embryonic breast development. Development *128*, 513-525.

Ford, A.M., Bennett, C.A., Healy, L.E., Towatari, M., Greaves, M.F., and Enver, T. (1996). Regulation of the myeloperoxidase enhancer binding proteins Pu1, C-EBP alpha, -beta, and -delta during granulocyte-lineage specification. Proc. Natl. Acad. Sci. U.S.A. *93*, 10838–10843.

Gallego, M.I., Binart, N., Robinson, G.W., Okagaki, R., Coschigano, K.T., Perry, J., Kopchick, J.J., Oka, T., Kelly, P.A., and Hennighausen, L. (2001). Prolactin, growth hormone, and epidermal growth factor activate Stat5 in different compartments of mammary tissue and exert different and overlapping developmental effects. Developmental Biology *229*, 163–175.

Gastaldi, S., Sassi, F., Accornero, P., Torti, D., Galimi, F., Migliardi, G., Molyneux, G., Perera, T., Comoglio, P.M., Boccaccio, C., et al. (2012). Met signaling regulates growth, repopulating potential and basal cell-fate commitment of mammary luminal progenitors: implications for basal-like breast cancer. Oncogene, 1-13.

Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., et al. (2004). Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. *5*, R80.

Gigliotti, A.P., and DeWille, J.W. (1998). Lactation status influences expression of CCAAT/enhancer binding protein isoform mRNA in the mouse mammary gland. J. Cell. Physiol. *174*, 232–239.

Gimpl, G., and Fahrenholz, F. (2001). The oxytocin receptor system: structure, function, and regulation. Physiol. Rev. *81*, 629–683.

Goldhar, A.S., Duan, R., Ginsburg, E., and Vonderhaar, B.K. (2011). Progesterone induces expression of the prolactin receptor gene through cooperative action of Sp1 and C/EBP. Molecular and Cellular Endocrinology *335*, 148–157.

Gomis, R.R., Alarcón, C., Nadal, C., Van Poznak, C., and Massagué, J. (2006). C/EBPbeta at the core of the TGFbeta cytostatic response and its evasion in metastatic breast cancer cells. Cancer Cell *10*, 203–214.

Goodell, M.A., Brose, K., Paradis, G., Conner, A.S., and Mulligan, R.C. (1996). Isolation and functional properties of murine hematopoietic stem cells that are replicating in vivo. J. Exp. Med. *183*, 1797–1806.

Graves, B.J., Johnson, P.F., and McKnight, S.L. (1986). Homologous recognition of a promoter domain common to the MSV LTR and the HSV tk gene. Cell *44*, 565–576.

Grimm, S.L., and Rosen, J.M. (2003). The Role of C/EBPbeta in Mammary Gland Development and Breast Cancer. J Mammary Gland Biol Neoplasia *8*, 191–204.

Gritli-Linde, A., Hallberg, K., Harfe, B.D., Reyahi, A., Kannius-Janson, M., Nilsson, J., Cobourne, M.T., Sharpe, P.T., McMahon, A.P., and Linde, A. (2007). Abnormal hair development and apparent follicular transformation to mammary gland in the absence of hedgehog signaling. Dev. Cell *12*, 99–112.

Guenette, R.S., Corbeil, H.B., Léger, J., Wong, K., Mézl, V., Mooibroek, M., and Tenniswood, M. (1994). Induction of gene expression during involution of the lactating mammary gland of the rat. J. Mol. Endocrinol. *12*, 47–60.

Guo, S., Cichy, S.B., He, X., Yang, Q., Ragland, M., Ghosh, A.K., Johnson, P.F., and Unterman, T.G. (2001). Insulin suppresses transactivation by CAAT/enhancer-binding proteins beta (C/EBPbeta). J. Biol. Chem. *276*, 8516–8523.

Haaksma, C.J., Schwartz, R.J., and Tomasek, J.J. (2011). Myoepithelial Cell Contraction and Milk Ejection Are Impaired in Mammary Glands of Mice Lacking Smooth Muscle Alpha-Actin. Biology of Reproduction *85*, 13–21.

Halbritter, F., Vaidya, H.J., and Tomlinson, S.R. (2012). GeneProf: analysis of high-throughput sequencing experiments. Nature Methods *9*, 7–8.

Han, Y., Watling, D., Rogers, N.C., and Stark, G.R. (1997). JAK2 and STAT5, but not JAK1 and STAT1, are required for prolactin-induced beta-lactoglobulin transcription. Molecular Endocrinology *11*, 1180–1188.

Harris, J., Stanford, P.M., Sutherland, K., Oakes, S.R., Naylor, M.J., Robertson, F.G., Blazek, K.D., Kazlauskas, M., Hilton, H.N., Wittlin, S., et al. (2006). Socs2 and Elf5 mediate prolactin-induced mammary gland development. Molecular Endocrinology *20*, 1177–1187.

Haslam, R.J., Koide, H.B., and Hemmings, B.A. (1993). Pleckstrin domain homology. Nature *363*, 309–310.

Haslam, S.Z., and Shyamala, G. (1979). Effect of oestradiol on progesterone receptors in normal mammary glands and its relationship with lactation. Biochem. J. *182*, 127–131.

Hatsell, S.J., and Cowin, P. (2006). Gli3-mediated repression of Hedgehog targets is required for normal mammary development. Development *133*, 3661–3670.

Heckman, B.M., Chakravarty, G., Vargo-Gogola, T., Gonzales-Rimbau, M., Hadsell, D.L., Lee, A.V., Settleman, J., and Rosen, J.M. (2007). Crosstalk between the p190-B RhoGAP and IGF signaling pathways is required for embryonic mammary bud development. Developmental Biology *309*, 137–149.

Hennighausen, L., and Robinson, G.W. (2005). Information networks in the mammary gland. Nat Rev Mol Cell Biol *6*, 715–725.

Hens, J., Dann, P., Zhang, J., and Harris, S. (2007). BMP4 and PTHrP interact to stimulate ductal outgrowth during embryonic mammary development and to inhibit hair follicle induction. Development *134*, 1221-1230.

Heuberger, B., Fitzka, I., Wasner, G., and Kratochwil, K. (1982). Induction of androgen receptor formation by epithelium-mesenchyme interaction in embryonic mouse mammary gland. Proc. Natl. Acad. Sci. U.S.A. *79*, 2957–2961.

Hoffmann, P.R., deCathelineau, A.M., Ogden, C.A., Leverrier, Y., Bratton, D.L., Daleke, D.L., Ridley, A.J., Fadok, V.A., and Henson, P.M. (2001). Phosphatidylserine (PS) induces PS receptor-mediated macropinocytosis and promotes clearance of apoptotic cells. The Journal of Cell Biology *155*, 649–659.

Howard, B., Panchal, H., McCarthy, A., and Ashworth, A. (2005). Identification of the scaramanga gene implicates Neuregulin3 in mammary gland specification. Genes & Development *19*, 2078–2090.

Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2008). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc *4*, 44–57.

Humphreys, R.C., Bierie, B., Zhao, L., Raz, R., Levy, D., and Hennighausen, L. (2002). Deletion of Stat3 blocks mammary gland involution and extends functional competence of the secretory epithelium in the absence of lactogenic stimuli. Endocrinology *143*, 3641–3650.

Humphreys, R.C., Krajewska, M., Krnacik, S., Jaeger, R., Weiher, H., Krajewski, S., Reed, J.C., and Rosen, J.M. (1996). Apoptosis in the terminal endbud of the murine mammary gland: a mechanism of ductal morphogenesis. Development *122*, 4013-4022.

Hung, K.-W., Kumar, T.K.S., Kathir, K.M., Xu, P., Ni, F., Ji, H.-H., Chen, M.-C., Yang, C.-C., Lin, F.-P., Chiu, I.-M., et al. (2005). Solution structure of the ligand binding domain of the fibroblast growth factor receptor: role of heparin in the activation of the receptor. Biochemistry *44*, 15787–15798.

Hurtado, A., Holmes, K.A., Ross-Innes, C.S., Schmidt, D., and Carroll, J.S. (2011). FOXA1 is a key determinant of estrogen receptor function and endocrine response. Nat. Genet. *43*, 27–33.

Janssen, B.J.C., Robinson, R.A., Pérez-Brangulí, F., Bell, C.H., Mitchell, K.J., Siebold, C., and Jones, E.Y. (2010). Structural basis of semaphorin-plexin signalling. Nature *467*, 1118–1122.

Jesselsohn, R., Brown, N.E., Arendt, L., Klebba, I., Hu, M.G., Kuperwasser, C., and Hinds, P.W. (2010). Cyclin D1 Kinase Activity Is Required for the Self-Renewal of Mammary Stem and Progenitor Cells that Are Targets of MMTV-ErbB2 Tumorigenesis. Cancer Cell *17*, 65–76.

Jones, P.H., Harper, S., and Watt, F.M. (1995). Stem cell patterning and fate in human epidermis. Cell *80*, 83–93.

Joshi, P.A., Di Grappa, M.A., and Khokha, R. (2012). Active allies: hormones, stem cells and the niche in adult mammopoiesis. Trends in Endocrinology & Metabolism *23*, 299–309.

Karaya, K., Mori, S., Kimoto, H., Shima, Y., Tsuji, Y., Kurooka, H., Akira, S., and Yokota, Y. (2005). Regulation of Id2 expression by CCAAT/enhancer binding protein β. Nucleic Acids Res. *33*, 1924-1934.

Kastner, P., Krust, A., Turcotte, B., Stropp, U., Tora, L., Gronemeyer, H., and Chambon, P. (1990). Two distinct estrogen-regulated promoters generate transcripts encoding the two functionally different human progesterone receptor forms A and B. Embo J. *9*, 1603-1614.

Kelleher, S.L., and Lönnerdal, B. (2001). Immunological activities associated with milk. Adv Nutr Res *10*, 39–65.

Kendrick, H., Regan, J.L., Magnay, F.-A., Grigoriadis, A., Mitsopoulos, C., Zvelebil, M., and Smalley, M.J. (2008). Transcriptome analysis of mammary epithelial subpopulations identifies novel determinants of lineage commitment and cell fate. BMC Genomics *9*, 591.

Kim, H.J., Yoon, M.-J., Lee, J., Penninger, J.M., and Kong, Y.-Y. (2002). Osteoprotegerin ligand induces beta-casein gene expression through the transcription factor CCAAT/enhancer-binding protein beta. J. Biol. Chem. *277*, 5339–5344.

Kleinberg, D.L., Ruan, W., Catanese, V., Newman, C.B., and Feldman, M. (1990). Non-lactogenic effects of growth hormone on growth and insulin-like growth factor-I messenger ribonucleic acid of rat mammary gland. Endocrinology *126*, 3274–3276.

Kordon, E.C., and Smith, G.H. (1998). An entire functional mammary gland may comprise the progeny from a single cell. Development *125*, 1921–1930.

Kouros-Mehr, H., Slorach, E.M., Sternlicht, M.D., and Werb, Z. (2006). GATA-3 Maintains the Differentiation of the Luminal Cell Fate in the Mammary Gland. Cell *127*, 1041–1055.

Kowenz-Leutz, E., and Leutz, A. (1999). A C/EBP beta isoform recruits the SWI/SNF complex to activate myeloid genes. Mol. Cell *4*, 735–743.

Kowenz-Leutz, E., Twamley, G., Ansieau, S., and Leutz, A. (1994). Novel mechanism of C/EBP beta (NF-M) transcriptional control: activation through derepression. Genes & Development *8*, 2781–2791.

Kritikou, E.A., Sharkey, A., Abell, K., Came, P.J., Anderson, E., Clarkson, R.W.E., and Watson, C.J. (2003). A dual, non-redundant, role for LIF as a regulator of development and STAT3-mediated cell death in mammary gland. Development *130*, 3459–3468.

Kruger, R.P., Aurandt, J., and Guan, K.-L. (2005). Semaphorins command cells to move. Nat Rev Mol Cell Biol *6*, 789–800.

Kuhn, N.J. (1969). Progesterone withdrawal as the lactogenic trigger in the rat. J Endocrinol *44*, 39–54.

LaMarca, H.L., and Rosen, J.M. (2007). Estrogen regulation of mammary gland development and breast cancer: amphiregulin takes center stage. Breast Cancer Res *9*, 304.

LaMarca, H.L., and Rosen, J.M. (2008). Hormones and Mammary Cell Fate-What Will I Become When I Grow Up? Endocrinology *149*, 4317–4321.

LaMarca, H.L., Visbal, A.P., Creighton, C.J., Liu, H., Zhang, Y., Behbod, F., and Rosen, J.M. (2010). CCAAT/enhancer binding protein beta regulates stem cell activity and specifies luminal cell fate in the mammary gland. Stem Cells *28*, 535-544.

Lamb, J., Ramaswamy, S., Ford, H.L., Contreras, B., Martinez, R.V., Kittrell, F.S., Zahnow, C.A., Patterson, N., Golub, T.R., and Ewen, M.E. (2003). A mechanism of cyclin D1 action encoded in the patterns of gene expression in human cancer. Cell *114*, 323–334.

Landschulz, W.H., Johnson, P.F., and McKnight, S.L. (1989). The DNA binding domain of the rat liver nuclear protein C/EBP is bipartite. Science *243*, 1681–1688.

Larive, R.M., Abad, A., Cardaba, C.M., Hernández, T., Cañamero, M., de Álava, E., Santos, E., Alarcón, B., and Bustelo, X.R. (2012). The Ras-like protein R-Ras2/TC21 is important for proper mammary gland development. Mol. Biol. Cell *23*, 2373–2387.

LeClair, K.P., Blanar, M.A., and Sharp, P.A. (1992). The p50 subunit of NF-KB associates with the NF-IL6 transcription factor. Proc. Natl. Acad. Sci. U.S.A. *89*, 1-5.

Lekstrom-Himes, J., and Xanthopoulos, K.G. (1998). Biological role of the CCAAT/enhancer-binding protein family of transcription factors. Journal of Biological Chemistry *273*, 28545–28548.

Li, A., Simmons, P.J., and Kaur, P. (1998). Identification and isolation of candidate human keratinocyte stem cells based on cell surface phenotype. Proc. Natl. Acad. Sci. U.S.A. *95*, 3902–3907.

Li, F., Strange, R., Friis, R.R., Djonov, V., Altermatt, H.J., Saurer, S., Niemann, H., and Andres, A.C. (1994). Expression of stromelysin-1 and TIMP-1 in the involuting mammary gland and in early invasive tumors of the mouse. Int. J. Cancer *59*, 560–568.

Li, M., Liu, X., Robinson, G., Bar-Peled, U., Wagner, K.U., Young, W.S., Hennighausen, L., and Furth, P.A. (1997). Mammary-derived signals activate programmed cell death during the first stage of mammary gland involution. Proc. Natl. Acad. Sci. U.S.A. *94*, 3425–3430.

Li, N., Zhang, Y., Naylor, M.J., Schatzmann, F., Maurer, F., Wintermantel, T., Schuetz, G., Mueller, U., Streuli, C.H., and Hynes, N.E. (2005). Beta1 integrins regulate mammary gland proliferation and maintain the integrity of mammary alveoli. Embo J. *24*, 1942–1953.

Li, W., Ferguson, B.J., Khaled, W.T., Tevendale, M., Stingl, J., Poli, V., Rich, T., Salomoni, P., and Watson, C.J. (2009). PML depletion disrupts normal mammary gland development and skews the composition of the mammary luminal cell progenitor pool. Proc. Natl. Acad. Sci. U.S.A. *106*, 4725.

Linhart, H.G., Ishimura-Oka, K., DeMayo, F., Kibe, T., Repka, D., Poindexter, B., Bick, R.J., and Darlington, G.J. (2001). C/EBPalpha is required for differentiation of white, but not brown, adipose tissue. Proc. Natl. Acad. Sci. U.S.A. *98*, 12532–12537.

Liu, B.Y., McDermott, S.P., Khwaja, S.S., and Alexander, C.M. (2004). The transforming activity of Wnt effectors correlates with their ability to induce the accumulation of mammary progenitor cells. Proc. Natl. Acad. Sci. U.S.A. *101*, 4158–4163.

Liu, X., Robinson, G.W., Wagner, K.U., Garrett, L., Wynshaw-Boris, A., and Hennighausen, L. (1997). Stat5a is mandatory for adult mammary gland development and lactogenesis. Genes & Development *11*, 179–186.

Long, W., Wagner, K.-U., Lloyd, K.C.K., Binart, N., Shillingford, J.M., Hennighausen, L., and Jones, F.E. (2003). Impaired differentiation and lactational failure of Erbb4-deficient mammary glands identify ERBB4 as an obligate mediator of STAT5. Development *130*, 5257–5268.

Lopez, J.I., Camenisch, T.D., Stevens, M.V., Sands, B.J., McDonald, J., and Schroeder, J.A. (2005). CD44 attenuates metastatic invasion during breast cancer progression. Cancer Research *65*, 6755–6763.

Louderbough, J.M.V., Brown, J.A., Nagle, R.B., and Schroeder, J.A. (2011). CD44 Promotes Epithelial Mammary Gland Development and Exhibits Altered Localization during Cancer Progression. Genes Cancer *2*, 771–781.

Lupien, M., Eeckhoute, J., Meyer, C.A., Wang, Q., Zhang, Y., Li, W., Carroll, J.S., Liu, X.S., and Brown, M. (2008). FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. Cell *132*, 958–970.

Lydon, J.P., DeMayo, F.J., Funk, C.R., Mani, S.K., Hughes, A.R., Montgomery, C.A., Shyamala, G., Conneely, O.M., and O'Malley, B.W. (1995). Mice lacking progesterone receptor exhibit pleiotropic reproductive abnormalities. Genes & Development *9*, 2266–2278.

Lyons, W., Li, C., and aJohnson, R. (1958). The hormonal control of mammary growth and lactation. Recent Prog. Horm. Res. *14*, 219–48.

Mailleux, A.A., Spencer-Dene, B., Dillon, C., Ndiaye, D., Savona-Baron, C., Itoh, N., Kato, S., Dickson, C., Thiery, J.P., and Bellusci, S. (2002). Role of FGF10/FGFR2b signaling during mammary gland development in the mouse embryo. Development *129*, 53–60.

Mallepell, S., Krust, A., Chambon, P., and Brisken, C. (2006). Paracrine signaling through the epithelial estrogen receptor alpha is required for proliferation and morphogenesis in the mammary gland. Proc. Natl. Acad. Sci. U.S.A. *103*, 2196–2201.

Mancini, E., Sanjuan-Pla, A., Luciani, L., Moore, S., Grover, A., Zay, A., Rasmussen, K.D., Luc, S., Bilbao, D., O'Carroll, D., et al. (2011). FOG-1 and GATA-1 act sequentially to specify definitive megakaryocytic and erythroid progenitors. EMBO J. *31*, 351–365.

Mannstadt, M., Jüppner, H., and Gardella, T.J. (1999). Receptors for PTH and PTHrP: their biological importance and functional properties. Am. J. Physiol. *277*, F665–F675.

Maroulakou, I.G., Oemler, W., Naber, S.P., Klebba, I., Kuperwasser, C., and Tsichlis, P.N. (2008). Distinct roles of the three Akt isoforms in lactogenic differentiation and involution. J. Cell. Physiol. *217*, 468–477.

Marti, A., Lazar, H., Ritter, P., and Jaggi, R. (1999). Transcription factor activities and gene expression during mouse mammary gland involution. J Mammary Gland Biol Neoplasia *4*, 145–152.

Mather, I.H., and Keenan, T.W. (1998). Origin and secretion of milk lipids. J Mammary Gland Biol Neoplasia *3*, 259–273.

Matise, L.A., Palmer, T.D., Ashby, W.J., Nashabi, A., Chytil, A., Aakre, M., Pickup, M.W., Gorska, A.E., Zijlstra, A., and Moses, H.L. (2012). Lack of transforming growth factor-β signaling promotes collective cancer cell invasion through tumor-stromal crosstalk. Breast Cancer Res *14*, R98.

Matsuzawa, A., Nakano, H., Yoshimoto, T., and Sayama, K. (1995). Biology of mouse mammary tumor virus (MMTV). Cancer Lett. *90*, 3–11.

Mink, S., Haenig, B., and Klempnauer, K.H. (1997). Interaction and functional collaboration of p300 and C/EBPbeta. Molecular and Cellular Biology *17*, 6609–6617.

Mo, X., Kowenz-Leutz, E., Xu, H., and Leutz, A. (2004). Ras induces mediator complex exchange on C/EBP beta. Mol. Cell *13*, 241–250.

Molenaar, A.J., Kuys, Y.M., Davis, S.R., Wilkins, R.J., Mead, P.E., and Tweedie, J.W. (1996). Elevation of lactoferrin gene expression in developing, ductal, resting, and regressing parenchymal epithelium of the ruminant mammary gland. J. Dairy Sci. *79*, 1198–1208.

Monaghan, P., and Moss, D. (1996). Connexin expression and gap junctions in the mammary gland. Cell Biol. Int. *20*, 121–125.

Monks, J., Rosner, D., Geske, F.J., Lehman, L., Hanson, L., Neville, M.C., and Fadok, V.A. (2005). Epithelial cells as phagocytes: apoptotic epithelial cells are engulfed by mammary alveolar epithelial cells and repress inflammatory mediator release. Cell Death Differ. *12*, 107–114.

Monks, J., Smith-Steinhart, C., Kruk, E.R., Fadok, V.A., and Henson, P.M. (2008). Epithelial cells remove apoptotic epithelial cells during post-lactation involution of the mouse mammary gland. Biology of Reproduction *78*, 586–594.

Moreira, J.M.A., Cabezón, T., Gromova, I., Gromov, P., Timmermans-Wielenga, V., Machado, I., Llombart-Bosch, A., Kroman, N., Rank, F., and Celis, J.E. (2010). Tissue proteomics of the human mammary gland: towards an abridged definition of the molecular phenotypes underlying epithelial normalcy. Mol Oncol *4*, 539–561.

Mori, S., Nishikawa, S.I., and Yokota, Y. (2000). Lactation defect in mice lacking the helix-loop-helix inhibitor Id2. EMBO J. *19*, 5772–5781.

Morrison, B.L., Jose, C.C., and Cutler, M.L. (2010). Connective Tissue Growth Factor (CTGF/CCN2) enhances lactogenic differentiation of mammary epithelial cells via integrin-mediated cell adhesion. BMC Cell Biol. *11*, 35.

Mulac-Jericevic, B., Lydon, J.P., DeMayo, F.J., and Conneely, O.M. (2003). Defective mammary gland morphogenesis in mice lacking the progesterone receptor B isoform. Proc. Natl. Acad. Sci. U.S.A. *100*, 9744–9749.

Mulac-Jericevic, B., Mullinax, R.A., DeMayo, F.J., Lydon, J.P., and Conneely, O.M. (2000). Subgroup of Reproductive Functions of Progesterone Mediated by Progesterone Receptor-B Isoform. Science Signaling *289*, 1751.

Murphy, L.J., Bell, G.I., and Friesen, H.G. (1987). Growth hormone stimulates sequential induction of c-myc and insulin-like growth factor I expression in vivo. Endocrinology *120*, 1806–1812.

Naylor, M., Li, N., Cheung, J., and Lowe, E. (2005). Ablation of β1 integrin in mammary epithelium reveals a key role for integrin in glandular morphogenesis and differentiation. The Journal of Cell Biology *171*, 717-728.

Nerlov, C. (2007). The C/EBP family of transcription factors: a paradigm for interaction between gene expression and proliferation control. Trends in Cell Biology *17*, 318-324.

Neville, M.C., McFadden, T.B., and Forsyth, I. (2002). Hormonal regulation of mammary differentiation and milk secretion. J Mammary Gland Biol Neoplasia *7*, 49–66.

Ng, S.T., Zhou, J., Adesanya, O.O., Wang, J., LeRoith, D., and Bondy, C.A. (1997). Growth hormone treatment induces mammary gland hyperplasia in aging primates. Nat Med *3*, 1141–1144.

Nguyen, A.V., and Pollard, J.W. (2000). Transforming growth factor beta3 induces cell death during the first stage of mammary gland involution. Development *127*, 3107–3118.

Nguyen, D.A., Parlow, A.F., and Neville, M.C. (2001). Hormonal regulation of tight junction closure in the mouse mammary epithelium during the transition from pregnancy to lactation. J Endocrinol *170*, 347–356.

Niehof, M. (2001). Autoregulation enables different pathways to control CCAAT/enhancer binding protein β (C/EBPβ) transcription. Journal of Molecular Biology *309*, 855–868.

Niehof, M., Manns, M.P., and Trautwein, C. (1997). CREB controls LAP/C/EBP beta transcription. Molecular and Cellular Biology *17*, 3600–3613.

Nishimori, K., Young, L.J., Guo, Q., Wang, Z., Insel, T.R., and Matzuk, M.M. (1996). Oxytocin is required for nursing but is not essential for parturition or reproductive behavior. Proc. Natl. Acad. Sci. U.S.A. *93*, 11699–11704.

Nishio, Y., Isshiki, H., Kishimoto, T., and Akira, S. (1993). A nuclear factor for interleukin-6 expression (NF-IL6) and the glucocorticoid receptor synergistically activate transcription of the rat alpha 1-acid glycoprotein gene via direct protein-protein interaction. Molecular and Cellular Biology *13*, 1854–1862.

Novitskaya, V., Romanska, H., Dawoud, M., Jones, J.L., and Berditchevski, F. (2010). Tetraspanin CD151 regulates growth of mammary epithelial cells in three-dimensional extracellular matrix: implication for mammary ductal carcinoma in situ. Cancer Research *70*, 4698–4708.

O'Brien, J., Martinson, H., Durand-Rougely, C., and Schedin, P. (2012). Macrophages are crucial for epithelial cell death and adipocyte repopulation during mammary gland involution. Development *139*, 269–275.

Oakes, S.R., Hilton, H.N., and Ormandy, C.J. (2006). The alveolar switch: coordinating the proliferative cues and cell fate decisions that drive the formation of lobuloalveoli from ductal epithelium. Breast Cancer Res *8*, 207-217.

Oakes, S.R., Naylor, M.J., Asselin-Labat, M.L., Blazek, K.D., Gardiner-Garden, M., Hilton, H.N., Kazlauskas, M., Pritchard, M.A., Chodosh, L.A., Pfeffer, P.L., et al. (2008). The Ets transcription factor Elf5 specifies mammary alveolar cell fate. Genes & Development *22*, 581–586.

Ormandy, C.J., Binart, N., and Kelly, P.A. (1997a). Mammary gland development in prolactin receptor knockout mice. J Mammary Gland Biol Neoplasia *2*, 355–364.

Ormandy, C.J., Camus, A., Barra, J., Damotte, D., Lucas, B., Buteau, H., Edery, M., Brousse, N., Babinet, C., Binart, N., et al. (1997b). Null mutation of the prolactin receptor gene produces multiple reproductive defects in the mouse. Genes & Development *11*, 167–178.

Ormandy, C.J., Naylor, M., Harris, J., Robertson, F., Horseman, N.D., Lindeman, G.J., Visvader, J., and Kelly, P.A. (2003). Investigation of the transcriptional changes underlying functional defects in the mammary glands of prolactin receptor knockout mice. Recent Prog. Horm. Res. *58*, 297–323.

Osada, S., Yamamoto, H., Nishihara, T., and Imagawa, M. (1996). DNA binding specificity of the CCAAT/enhancer-binding protein transcription factor family. Journal of Biological Chemistry *271*, 3891–3896.

Pavlovich, A.L., Boghaert, E., and Nelson, C.M. (2011). Mammary branch initiation and extension are inhibited by separate pathways downstream of TGFβ in culture. Experimental Cell Research *317*, 1872–1884.

Poli, V. (1998). The role of C/EBP isoforms in the control of inflammatory and native immunity functions. Journal of Biological Chemistry *273*, 29279–29282.

Porse, B.T., Pedersen, T.Å., Xu, X., Lindberg, B., Wewer, U.M., Friis-Hansen, L., and Nerlov, C. (2001). E2F repression by C/EBPα is required for adipogenesis and granulopoiesis in vivo. Cell *107*, 247–258.

Prager, E.M., and Wilson, A.C. (1988). Ancient origin of lactalbumin from lysozyme: analysis of DNA and amino acid sequences. J. Mol. Evol. *27*, 326–335.

Rainard, P., Riollet, C., Berthon, P., Cunha, P., Fromageau, A., Rossignol, C., and Gilbert, F.B. (2008). The chemokine CXCL3 is responsible for the constitutive chemotactic activity of bovine milk for neutrophils. Molecular Immunology *45*, 4020–4027.

Ramji, D.P., and Foka, P. (2002). CCAAT/enhancer-binding proteins: structure, function and regulation. Biochem. J. *365*, 561–575.

Raught, B., Gingras, A.C., James, A., Medina, D., Sonenberg, N., and Rosen, J.M. (1996). Expression of a translationally regulated, dominant-negative CCAAT/enhancer-binding protein β isoform and up-regulation of the eukaryotic translation initiation factor 2α are correlated with neoplastic transformation of mammary epithelial cells. Cancer Research *56*, 4382-4386.

Raught, B., Liao, W.S., and Rosen, J.M. (1995). Developmentally and hormonally regulated CCAAT/enhancer-binding protein isoforms influence beta-casein gene expression. Molecular Endocrinology *9*, 1223–1232.

Ray, B.K., and Ray, A. (1994). Expression of the gene encoding α1-acid glycoprotein in rabbit liver under acute-phase conditions involves induction and activation of β and δ CCAAT-enhancer-binding proteins. European Journal of Biochemistry *222*, 891-900.

Renard, C.A., Labalette, C., Armengol, C., Cougot, D., Wei, Y., Cairo, S., Pineau, P., Neuveut, C., de Reyniès, A., Dejean, A., et al. (2007). Tbx3 is a downstream target of the Wnt/beta-catenin pathway and a critical mediator of beta-catenin survival functions in liver cancer. Cancer Research *67*, 901–910.

Richert, M.M., Schwertfeger, K.L., Ryder, J.W., and Anderson, S.M. (2000). An atlas of mouse mammary gland development. J Mammary Gland Biol Neoplasia *5*, 227–241.

Robinson, G.W. (2007). Cooperation of signalling pathways in embryonic mammary gland development. Nature Reviews Genetics *8*, 963–972.

Robinson, G.W., Johnson, P.F., Hennighausen, L., and Sterneck, E. (1998). The C/EBPbeta transcription factor regulates epithelial cell proliferation and differentiation in the mammary gland. Genes & Development *12*, 1907–1916.

Robinson, G.W., Karpf, A.B.C., and Kratochwil, K. (1999). Regulation of mammary gland development by tissue interaction. J Mammary Gland Biol Neoplasia *4*, 9–19.

Robinson, G.W., McKnight, R.A., Smith, G.H., and Hennighausen, L. (1995). Mammary epithelial cells undergo secretory differentiation in cycling virgins but require pregnancy for the establishment of terminal differentiation. Development *121*, 2079-2090.

Ron, D., and Habener, J.F. (1992). CHOP, a novel developmentally regulated nuclear protein that dimerizes with transcription factors C/EBP and LAP and functions as a dominant-negative inhibitor of gene transcription. Genes & Development *6*, 439–453.

Ruan, W., and Kleinberg, D.L. (1999). Insulin-like growth factor I is essential for terminal end bud formation and ductal morphogenesis during mammary development. Endocrinology *140*, 5075–5081.

Ruan, W., Newman, C.B., and Kleinberg, D.L. (1992). Intact and amino-terminally shortened forms of insulin-like growth factor I induce mammary gland differentiation and development. Proc. Natl. Acad. Sci. U.S.A. *89*, 10872–10876.

Rudolph, M.C., McManaman, J.L., Hunter, L., Phang, T., and Neville, M.C. (2003). Functional development of the mammary gland: use of expression profiling and trajectory clustering to reveal changes in gene expression during pregnancy, lactation, and involution. J Mammary Gland Biol Neoplasia *8*, 287–307.

Ruffell, D., Mourkioti, F., Gambardella, A., Kirstetter, P., Lopez, R.G., Rosenthal, N., and Nerlov, C. (2009). A CREB-C/EBPβ cascade induces M2 macrophage-specific gene expression and promotes muscle injury repair. Proc. Natl. Acad. Sci. U.S.A. *106*, 17475–17480.

Sabatakos, G., Davies, G.E., Grosse, M., Cryer, A., and Ramji, D.P. (1998). Expression of the genes encoding CCAAT-enhancer binding protein isoforms in the mouse mammary gland during lactation and involution. Biochem. J. *334*, 205–210.

Sakakura, T. (1987). Mammary embryogenesis (New York: The mammary gland, development, regulation and function). New York: Plenum Pub Corp.

Sandmann, T., Jakobsen, J.S., and Furlong, E.E.M. (2007). ChIP-on-chip protocol for genome-wide analysis of transcription factor binding in Drosophila melanogaster embryos. Nat Protoc *1*, 2839–2855.

Schere-Levy, C., Buggiano, V., Quaglino, A., Gattelli, A., Cirio, M.C., Piazzon, I., Vanzulli, S., and Kordon, E.C. (2003). Leukemia inhibitory factor induces apoptosis of the mammary epithelial cells and participates in mouse mammary gland involution. Experimental Cell Research *282*, 35–47.

Schmidt, D., Wilson, M.D., Ballester, B., Schwalie, P.C., Brown, G.D., Marshall, A., Kutter, C., Watt, S., Martinez-Jimenez, C.P., Mackay, S., et al. (2010). Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. Science *328*, 1036–1040.

Schmidt, D., Wilson, M.D., Spyrou, C., Brown, G.D., Hadfield, J., and Odom, D.T. (2009). ChIP-seq: Using high-throughput sequencing to discover protein–DNA interactions. Methods *48*, 240–248.

Schwertfeger, K.L., Richert, M.M., and Anderson, S.M. (2001). Mammary gland involution is delayed by activated Akt in transgenic mice. Molecular Endocrinology *15*, 867–881.

Screpanti, I., Romani, L., Musiani, P., Modesti, A., Fattori, E., Lazzaro, D., Sellitto, C., Scarpa, S., Bellavia, D., Lattanzio, G., et al. (1995). Lymphoproliferative disorder and imbalanced T-helper response is C/EBPbeta-deficient mice. EMBO J. *14*, 1932–1941.

Seagroves, T.N., Krnacik, S., Raught, B., Gay, J., Burgess-Beusse, B., Darlington, G.J., and Rosen, J.M. (1998). C/EBPbeta , but not C/EBPalpha , is essential for ductal morphogenesis, lobuloalveolar proliferation, and functional differentiation in the mouse mammary gland. Genes & Development *12*, 1917–1928.

Seagroves, T.N., Lydon, J.P., Hovey, R.C., Vonderhaar, B.K., and Rosen, J.M. (2000). C/EBPbeta (CCAAT/enhancer binding protein) controls cell fate determination during mammary gland development. Molecular Endocrinology *14*, 359–368.

Sebastian, T., Malik, R., Thomas, S., Sage, J., and Johnson, P.F. (2005). C/EBPbeta cooperates with RB:E2F to implement RasV12-induced cellular senescence. EMBO J. *24*, 3301–3312.

Seymour, K.J., Roberts, L.E., Fini, M.A., Parmley, L.A., Oustitch, T.L., and Wright, R.M. (2006). Stress activation of mammary epithelial cell xanthine oxidoreductase is mediated by p38 MAPK and CCAAT/enhancer-binding protein-beta. Journal of Biological Chemistry *281*, 8545–8558.

Shackleton, M., Vaillant, F., Simpson, K.J., Stingl, J., Smyth, G.K., Asselin-Labat, M.-L., Wu, L., Lindeman, G.J., and Visvader, J.E. (2006). Generation of a functional mammary gland from a single stem cell. Nature *439*, 84–88.

Shillingford, J.M., Miyoshi, K., Flagella, M., Shull, G.E., and Hennighausen, L. (2002a). Mouse mammary epithelial cells express the Na-K-Cl cotransporter, NKCC1: characterization, localization, and involvement in ductal development and morphogenesis. Molecular Endocrinology *16*, 1309–1321.

Shillingford, J.M., Miyoshi, K., Robinson, G.W., Grimm, S.L., Rosen, J.M., Neubauer, H., Pfeffer, K., and Hennighausen, L. (2002b). Jak2 is an essential tyrosine kinase involved in pregnancy-mediated development of mammary secretory epithelium. Molecular Endocrinology *16*, 563–570.

Sicinski, P., Donaher, J.L., Parker, S.B., Li, T., Fazeli, A., Gardner, H., Haslam, S.Z., Bronson, R.T., Elledge, S.J., and Weinberg, R.A. (1995). Cyclin D1 provides a link between development and oncogenesis in the retina and breast. Cell *82*, 621–630.

Sleeman, K.E., Kendrick, H., Ashworth, A., Isacke, C.M., and Smalley, M.J. (2006). CD24 staining of mouse mammary gland cells defines luminal epithelial, myoepithelial/basal and non-epithelial cells. Breast Cancer Res *8*, R7.

Sleeman, K.E., Kendrick, H., Robertson, D., Isacke, C.M., Ashworth, A., and Smalley, M.J. (2007). Dissociation of estrogen receptor expression and in vivo stem cell activity in the mammary gland. The Journal of Cell Biology *176*, 19–26.

Smith, G.H. (2005). Label-retaining epithelial cells in mouse mammary gland divide asymmetrically and retain their template DNA strands. Development *132*, 681–687.

Smith, G.H., and Medina, D. (1988). A morphologically distinct candidate for an epithelial stem cell in mouse mammary gland. J. Cell. Sci. *90*, 173–183.

Spangrude, G.J., Heimfeld, S., and Weissman, I.L. (1988). Purification and characterization of mouse hematopoietic stem cells. Science *241*, 58–62.

Srivastava, S., Matsuda, M., Hou, Z., Bailey, J.P., Kitazawa, R., Herbst, M.P., and Horseman, N.D. (2003). Receptor activator of NF-kappaB ligand induction via Jak2 and Stat5a in mammary epithelial cells. J. Biol. Chem. *278*, 46171–46178.

Stacey, A., Schnieke, A., Kerr, M., Scott, A., McKee, C., Cottingham, I., Binas, B., Wilde, C., and Colman, A. (1995). Lactation is disrupted by alpha-lactalbumin deficiency and can be restored by human alpha-lactalbumin gene replacement in mice. Proc. Natl. Acad. Sci. U.S.A. *92*, 2835–2839.

Sterneck, E., Tessarollo, L., and Johnson, P.F. (1997). An essential role for C/EBPbeta in female reproduction. Genes & Development *11*, 2153–2162.

Sternlicht, M.D. (2006). Key stages in mammary gland development: the cues that regulate ductal branching morphogenesis. Breast Cancer Res *8*, 201-212.

Stingl, J., Eirew, P., Ricketson, I., Shackleton, M., Vaillant, F., Choi, D., Li, H.I., and Eaves, C.J. (2006). Purification and unique properties of mammary epithelial stem cells. Nature *439*, 993-997.

Strange, R., Li, F., Saurer, S., Burkhardt, A., and Friis, R.R. (1992). Apoptotic cell death and tissue remodelling during mouse mammary gland involution. Development *115*, 49–58.

Streuli, C.H., and Bissell, M.J. (1990). Expression of extracellular matrix components is regulated by substratum. The Journal of Cell Biology *110*, 1405–1415.

Streuli, C.H., Edwards, G.M., Delcommenne, M., Whitelaw, C.B., Burdon, T.G., Schindler, C., and Watson, C.J. (1995). Stat5 as a target for regulation by extracellular matrix. J. Biol. Chem. *270*, 21639–21644.

Sun, P., Yuan, Y., Li, A., Li, B., and Dai, X. (2010). Cytokeratin expression during mouse embryonic and early postnatal mammary gland development. Histochem Cell Biol *133*, 213–221.

Takiguchi, M. (1998). The C/EBP family of transcription factors in the liver and other organs. Int J Exp Pathol *79*, 369–391.

Talhouk, R.S., Bissell, M.J., and Werb, Z. (1992). Coordinated expression of extracellular matrix-degrading proteinases and their inhibitors regulates mammary epithelial function during involution. The Journal of Cell Biology *118*, 1271–1282.

Tanaka, T., Akira, S., Yoshida, K., Umemoto, M., Yoneda, Y., Shirafuji, N., Fujiwara, H., Suematsu, S., Yoshida, N., and Kishimoto, T. (1995). Targeted disruption of the NF-IL6 gene discloses its essential role in bacteria killing and tumor cytotoxicity by macrophages. Cell *80*, 353–361.

Teglund, S., McKay, C., Schuetz, E., van Deursen, J.M., Stravopodis, D., Wang, D., Brown, M., Bodner, S., Grosveld, G., and Ihle, J.N. (1998). Stat5a and Stat5b proteins have essential and nonessential, or redundant, roles in cytokine responses. Cell *93*, 841–850.

Thangaraju, M., Rudelius, M., Bierie, B., Raffeld, M., Sharan, S., Hennighausen, L., Huang, A.-M., and Sterneck, E. (2005). C/EBPdelta is a crucial regulator of pro-apoptotic gene expression during mammary gland involution. Development *132*, 4675–4685.

Tkocz, D., Crawford, N.T., Buckley, N.E., Berry, F.B., Kennedy, R.D., Gorski, J.J., Harkin, D.P., and Mullan, P.B. (2011). BRCA1 and GATA3 corepress FOXC1 to inhibit the pathogenesis of basal-like breast cancers. Oncogene *31*, 3667–3678.

Tonner, E., Barber, M.C., Allan, G.J., Beattie, J., Webster, J., Whitelaw, C.B.A., and Flint, D.J. (2002). Insulin-like growth factor binding protein-5 (IGFBP-5) induces premature cell death in the mammary glands of transgenic mice. Development *129*, 4547–4557.

Trautwein, C., Caelles, C., van der Geer, P., Hunter, T., Karin, M., and Chojkier, M. (1993). Transactivation by NF-IL6/LAP is enhanced by phosphorylation of its activation domain. Nature *364*, 544–547.

Trautwein, C., van der Geer, P., Karin, M., Hunter, T., and Chojkier, M. (1994). Protein kinase A and C site-specific phosphorylations of LAP (NF-IL6) modulate its binding affinity to DNA recognition elements. J. Clin. Invest. *93*, 2554–2561.

Tu, Z.J., Kollander, R., and Kiang, D.T. (1998). Differential up-regulation of gap junction connexin 26 gene in mammary and uterine tissues: the role of Sp transcription factors. Molecular Endocrinology *12*, 1931–1938.

van Genderen, C., Okamura, R.M., Farinas, I., Quo, R.G., Parslow, T.G., Bruhn, L., and Grosschedl, R. (1994). Development of several organs that require inductive epithelial-mesenchymal interactions is impaired in LEF-1-deficient mice. Genes & Development *8*, 2691–2703.

Van Keymeulen, A., Rocha, A.S., Ousset, M., Beck, B., Bouvencourt, G., Rock, J., Sharma, N., Dekoninck, S., and Blanpain, C. (2011). Distinct stem cells contribute to mammary gland development and maintenance. Nature *479*, 189-193.

Vargo-Gogola, T., Heckman, B.M., Gunther, E.J., Chodosh, L.A., and Rosen, J.M. (2006). P190-B Rho GTPase-activating protein overexpression disrupts ductal morphogenesis and induces hyperplastic lesions in the developing mammary gland. Molecular Endocrinology *20*, 1391–1405.

Veltmaat, J.M., Mailleux, A.A., Thiery, J.P., and Bellusci, S. (2003). Mouse embryonic mammogenesis as a model for the molecular regulation of pattern formation. Differentiation *71*, 1–17.

Veltmaat, J.M., Relaix, F., Le, L.T., Kratochwil, K., Sala, F.G., Van Veelen, W., Rice, R., Spencer-Dene, B., Mailleux, A.A., Rice, D.P., et al. (2006). Gli3-mediated somitic Fgf10 expression gradients are required for the induction and patterning of mammary epithelium along the embryonic axes. Development *133*, 2325–2335.

Veltmaat, J.M., Van Veelen, W., Thiery, J.P., and Bellusci, S. (2004). Identification of the mammary line in mouse by Wnt10b expression. Developmental Dynamics *229*, 349–356.

Vinson, C.R., Sigler, P.B., and McKnight, S.L. (1989). Scissors-grip model for DNA recognition by a family of leucine zipper proteins. Science *246*, 911–916.

Vorbach, C., Capecchi, M.R., and Penninger, J.M. (2006). Evolution of the mammary gland from the innate immune system? Bioessays *28*, 606–616.

Vorbach, C., Harrison, R., and Capecchi, M.R. (2003). Xanthine oxidoreductase is central to the evolution and function of the innate immune system. Trends Immunol. *24*, 512–517.

Vorbach, C., Sciven, A., and Capecchi, M.R. (2002). The housekeeping gene xanthine oxidoreductase is necessary for milk fat droplet enveloping and secretion: gene sharing in the lactating mammary gland. Genes & Development *16*, 3223–3235.

Wagner, K.U., Boulanger, C.A., Henry, M.D., Sgagias, M., Hennighausen, L., and Smith, G.H. (2002). An adjunct mammary epithelial cell population in parous females: its role in functional adaptation and tissue renewal. Development *129*, 1377–1386.

Wakao, H., Gouilleux, F., and Groner, B. (1994). Mammary gland factor (MGF) is a novel member of the cytokine regulated transcription factor gene family and confers the prolactin response. EMBO J. *13*, 2182–2191.

Wang, W., Lv, N., Zhang, S., Shui, G., Qian, H., Zhang, J., Chen, Y., Ye, J., Xie, Y., Shen, Y., et al. (2012). Cidea is an essential transcriptional coactivator regulating mammary gland secretion of milk lipids. Nat Med *18*, 237–245.

Wang, W., Morrison, B., Galbaugh, T., Jose, C.C., Kenney, N., and Cutler, M.L. (2008). Glucocorticoid induced expression of connective tissue growth factor contributes to lactogenic differentiation of mouse mammary epithelial cells. J. Cell. Physiol. *214*, 38–46.

Watson, C.J., and Khaled, W.T. (2008). Mammary development in the embryo and adult: a journey of morphogenesis and commitment. Development *135*, 995–1003.

Wegner, M., Cao, Z., and Rosenfeld, M.G. (1992). Calcium-regulated phosphorylation within the leucine zipper of C/EBP beta. Science *256*, 370–373.

Welm, B.E., Tepera, S.B., Venezia, T., Graubert, T.A., Rosen, J.M., and Goodell, M.A. (2002). Sca-1pos Cells in the Mouse Mammary Gland Represent an Enriched Progenitor Cell Population. Developmental Biology *245*, 42–56.

Williams, S.C., Cantwell, C.A., and Johnson, P.F. (1991). A family of C/EBP-related proteins capable of forming covalently linked leucine zipper dimers in vitro. Genes & Development *5*, 1553–1567.

Wiseman, B.S., Sternlicht, M.D., Lund, L.R., Alexander, C.M., Mott, J., Bissell, M.J., Soloway, P., Itohara, S., and Werb, Z. (2003). Site-specific inductive and inhibitory activities of MMP-2 and MMP-3 orchestrate mammary gland branching morphogenesis. The Journal of Cell Biology *162*, 1123–1133.

Wysolmerski, J., Philbrick, W., and Dunbar, M. (1998). Rescue of the parathyroid hormone-related protein knockout mouse demonstrates that parathyroid hormone-related protein is essential for mammary gland development. Development *125*, 1285-1294.

Xing, J., Ginty, D.D., and Greenberg, M.E. (1996). Coupling of the RAS-MAPK pathway to gene activation by RSK2, a growth factor-regulated CREB kinase. Science *273*, 959–963.

Yamaji, D., Na, R., Feuermann, Y., Pechhold, S., Chen, W., Robinson, G.W., and Hennighausen, L. (2009). Development of mammary luminal progenitor cells is controlled by the transcription factor STAT5A. Genes & Development *23*, 2382–2387.

Young, L.J., Medina, D., DEOME, K.B., and Daniel, C.W. (1971). The influence of host and tissue age on life span and growth rate of serially transplanted mouse mammary gland. Exp. Gerontol. *6*, 49–56.

Zahnow, C.A., Cardiff, R.D., Laucirica, R., Medina, D., and Rosen, J.M. (2001). A role for CCAAT/enhancer binding protein beta-liver-enriched inhibitory protein in mammary epithelial cell proliferation. Cancer Research *61*, 261–269.

Zahnow, C.A., Younes, P., Laucirica, R., and Rosen, J.M. (1997). Overexpression of C/EBPbeta-LIP, a naturally occurring, dominant-negative transcription factor, in human breast cancer. J. Natl. Cancer Inst. *89*, 1887–1891.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137.

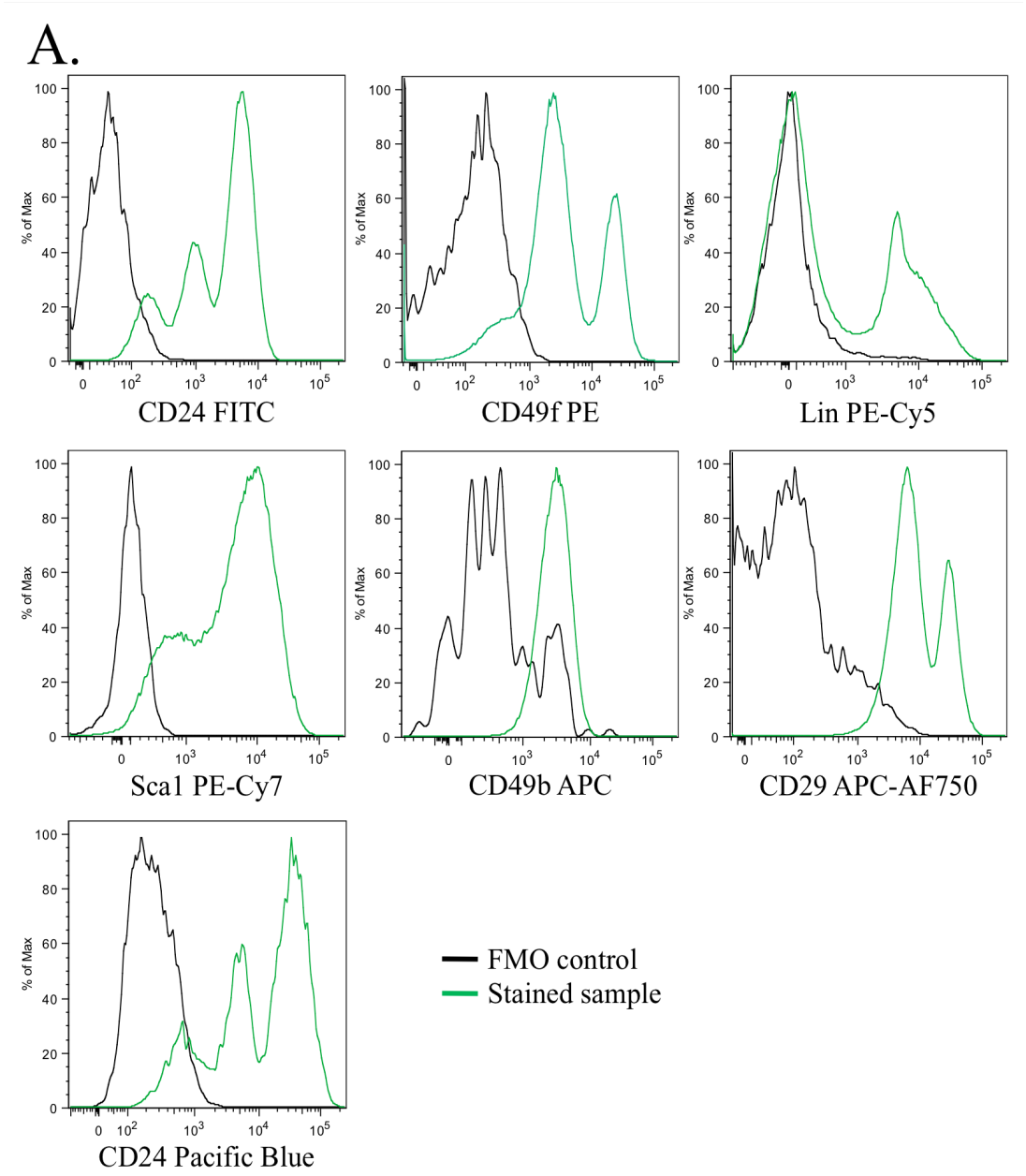
Zhou, J., Chehab, R., Tkalcevic, J., Naylor, M.J., Harris, J., Wilson, T.J., Tsao, S., Tellis, I., Zavarsek, S., Xu, D., et al. (2005). Elf5 is essential for early embryogenesis and mammary gland development during pregnancy and lactation. EMBO J. *24*, 635–644.

Zhu, S., Yoon, K., Sterneck, E., Johnson, P.F., and Smart, R.C. (2002). CCAAT/enhancer binding protein-beta is a mediator of keratinocyte survival and skin tumorigenesis involving oncogenic Ras signaling. Proc. Natl. Acad. Sci. U.S.A. *99*, 207–212.

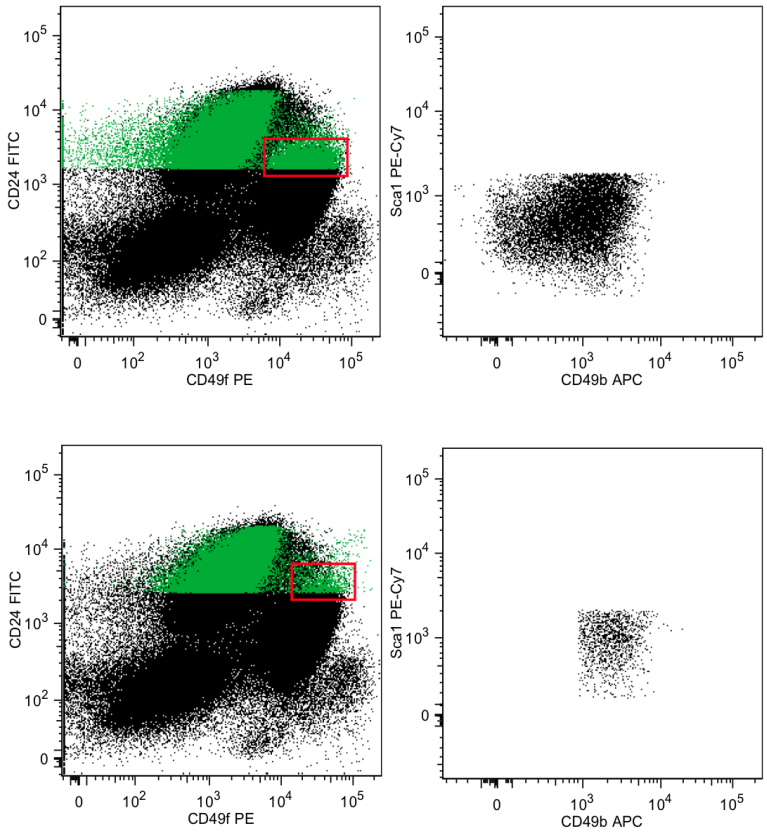
Zutter, M.M., Sun, H., and Santoro, S.A. (1998). Altered integrin expression and the malignant phenotype: the contribution of multiple integrated integrin receptors. J Mammary Gland Biol Neoplasia *3*, 191–200.

8 Appendices

Appendix 1. Accompaniment to Figure 2.1-2.5. (**A)** Representative fluoresence minus one controls used in flow cytometry. **(B)** Assessment of Sca1 expression of CD49f^{hi} cells in Sleeman's (top) and Li (bottom) gating strategies (Sleeman et al., 2007; Li et al., 2009).



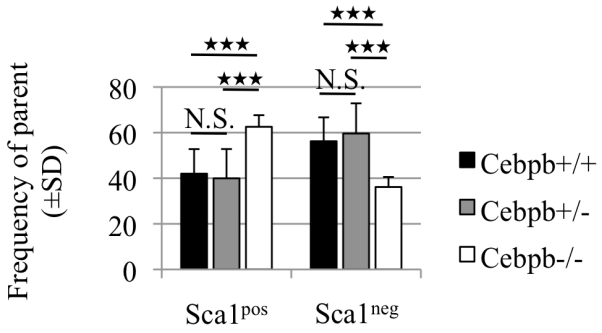
B.


CD24 FITC
CD49f PE
Sca1 PE-Cy7
CD49b APC
CD24 FITC
CD49f PE
Sca1 PE-Cy7
CD49b APC

Appendix 2. Accompaniment to Figure 2.8A. Complete functional annotation clustering analysis of differentially expressed genes in $Sca1^{pos}$ and $Sca1^{neg}$ luminal progenitor cells. Functional clusters generated by enriched genes in $Sca1^{pos}$ luminal progenitors. The transcript cluster IDs of differentially expressed genes (from Fig. 2.7) were analyzed using the DAVID Bioinformatics Resource using default parameters. Functional clusters were included in the figure if they had an enrichment score of ≥2.

Appendix 3. Accompaniment to Figure 2.8B. Complete functional annotation clustering analysis of differentially expressed genes in Sca1^{pos} and Sca1^{neg} luminal progenitor cells. Functional clusters generated by enriched genes in Sca1^{neg} luminal progenitors. The transcript cluster IDs of differentially expressed genes (from Fig. 2.7) were analyzed using the DAVID Bioinformatics Resource using default parameters. Functional clusters were included in the figure if they had an enrichment score of ≥2.

Appendix 4. Accompaniment to Figure 3.7. **C/EBPβ in luminal progenitor cell specification.** Bar graph and accompanying tables showing the frequency of the parent population ($Lin^{neg}CD29^{pos}CD24^{hi}CD49f^{lo}$) for $Sca1^{pos}$ cells and $Sca1^{neg}$ cells in *Cebpb*+/+ (n=4), *Cebpb*+/-, (n=4) and *Cebpb*-/- (n=10) mice. The values in the bar graph and accompanying tables show the mean frequency for each population. Error bars indicate standard deviation. Significance (p-value) was calculated using an unpaired, one-tailed Student's T-test. The level of significance is shown by stars in the graphs, and numerically in the tables. **Abbreviations:** SD, standard deviation; n, number; N.S., not significant.



% Parent ±SD	+/+ (n=4)	+/- (n=4)	-/- (n=10)
$Sca1^{pos}$	42.00 ±10.77	39.95 ±12.85	62.55 ±5.09

0.4075 N.S. 0.0002

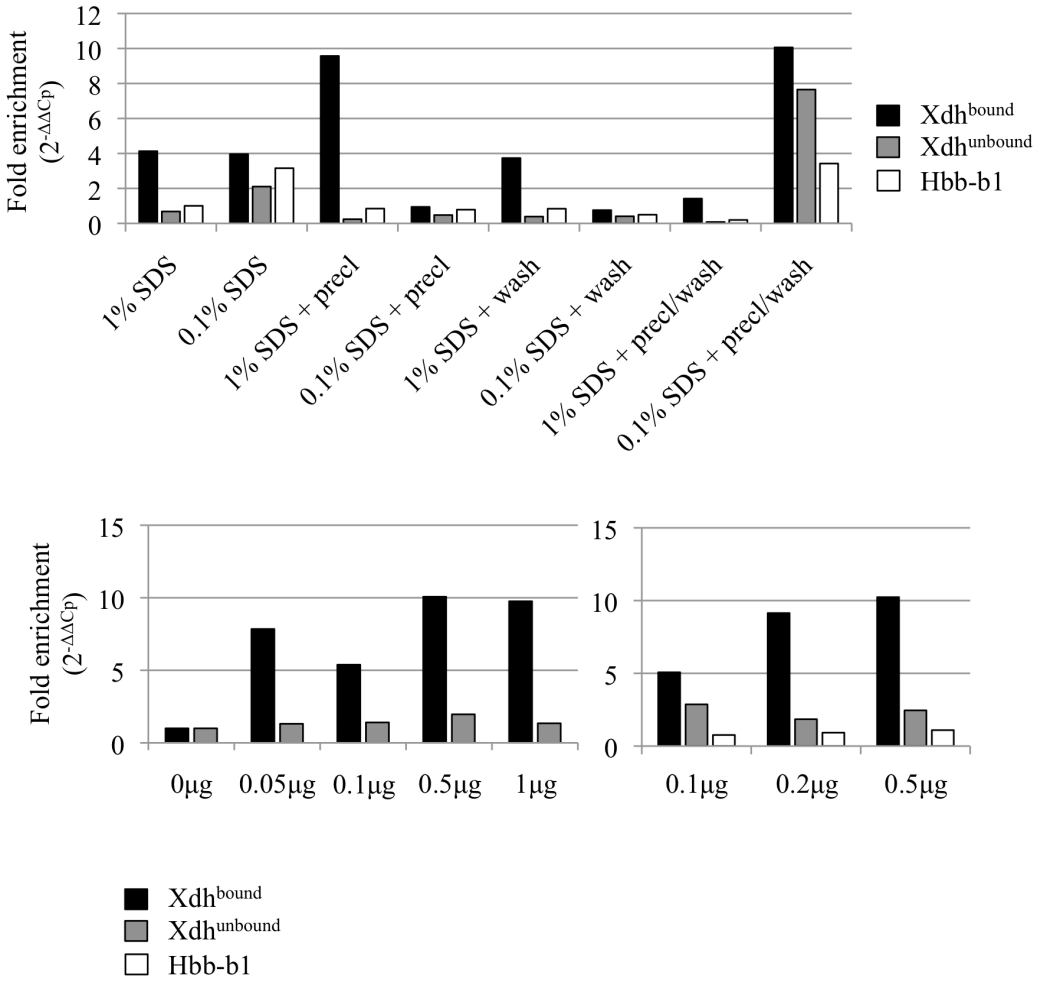
0.0002

% Parent ±SD	+/+ (n=4)	+/- (n=4)	-/- (n=10)
$Sca1^{neg}$	56.23 ±10.47	59.60 ±13.24	36.15 ±4.39

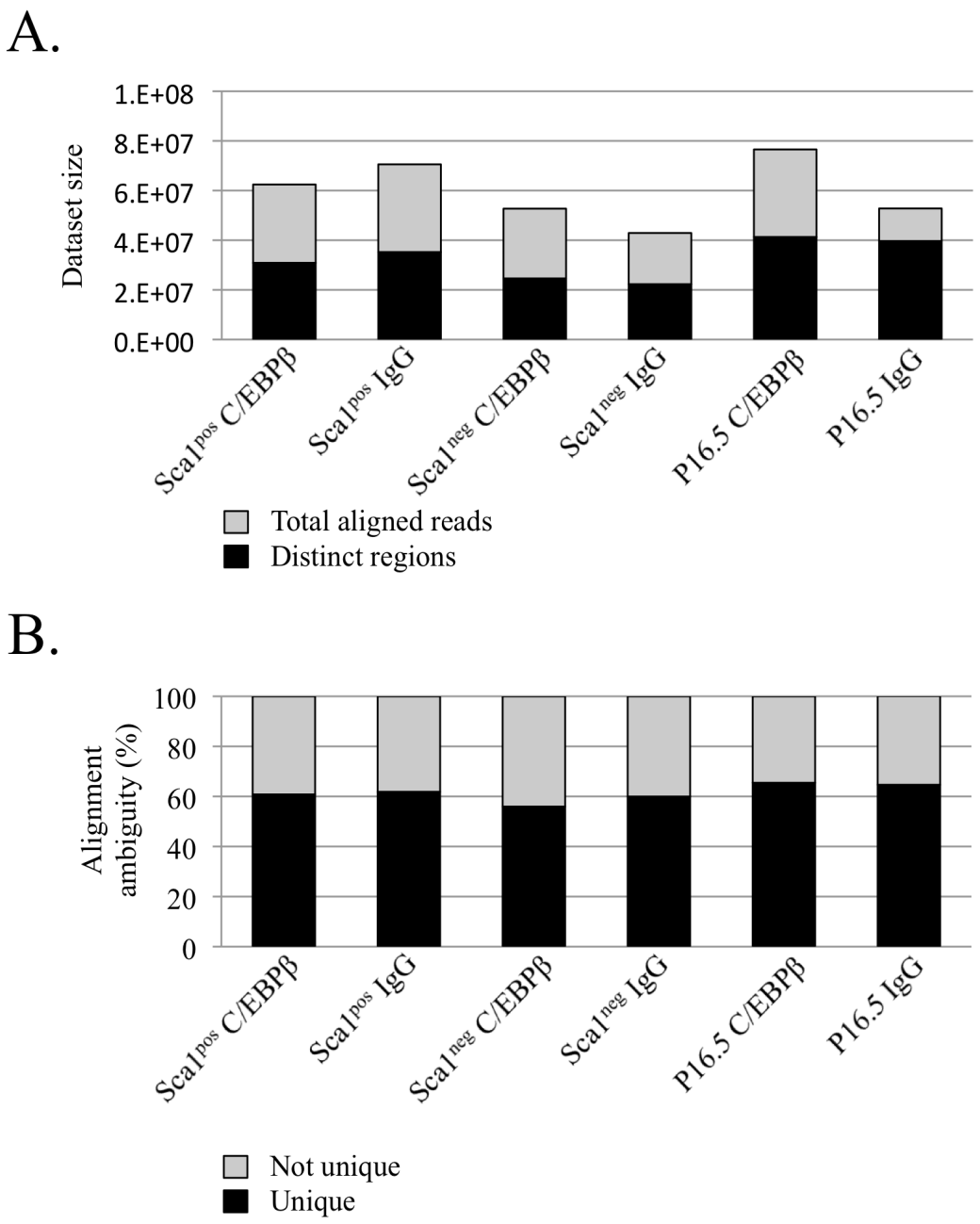
0.3515 N.S. 0.0001

0.0001

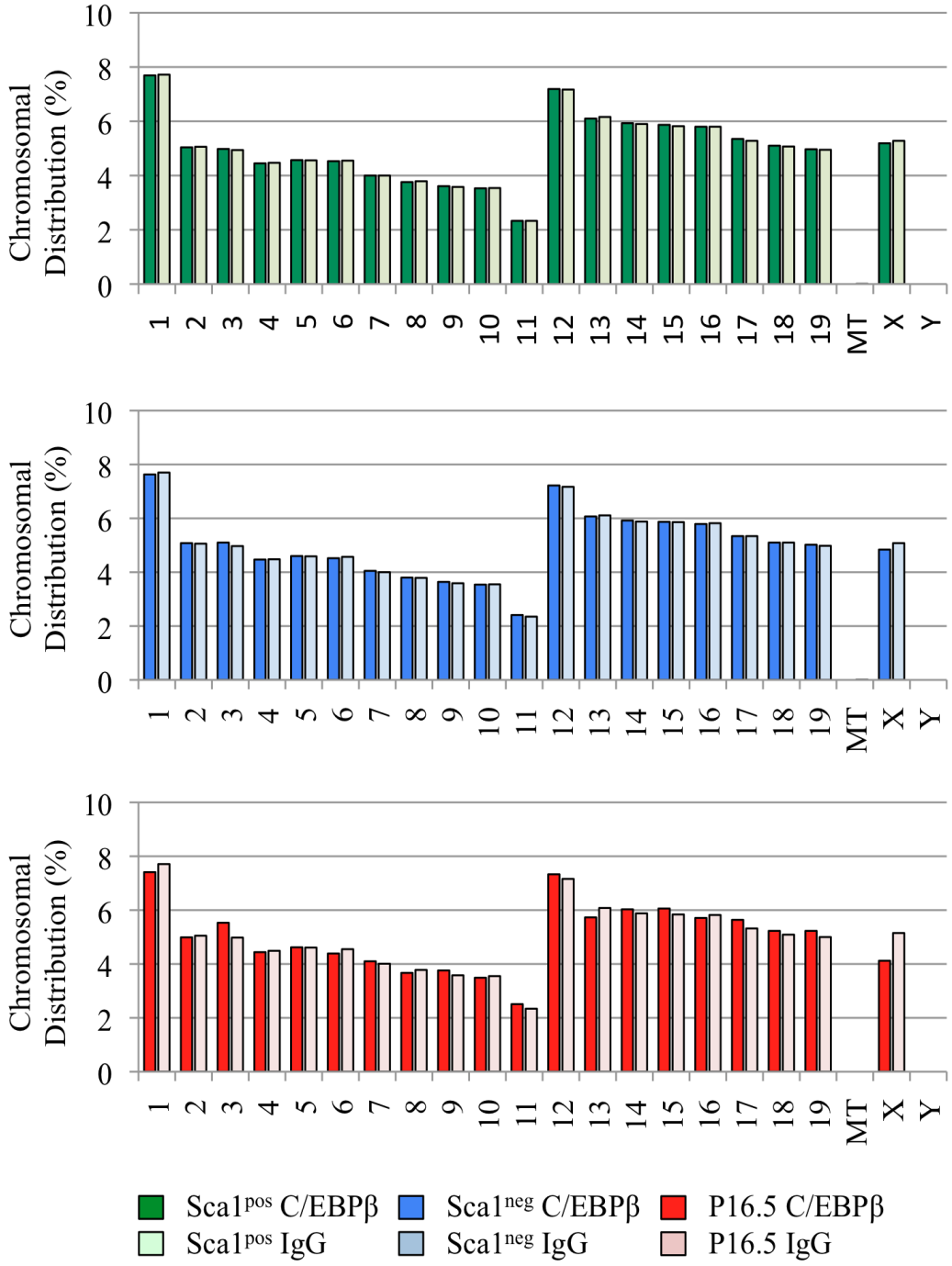
Appendix 5. Accompaniment to Figure 4.4. Development of the C/EBPβ chromatin immunoprecipitation (ChIP) method. Top bar graph depicts the optimization of the SDS concentration during sonication (1 % vs. 0.1 %), the inclusion of preclearing with Dynabeads Protein-A (± precl), and the inclusion of an additional washing step using a modified RIPA buffer that contains 500 mM NaCl and 0.5 % sodium deoxycholate (± wash). The optimal condition chosen was 1 % SDS concentration during sonication, and inclusion of a preclearing step, but not an additional wash step. Bottom bar graphs show the antibody titration of C19 (anti-C/EBPβ) (2 experiments). The optimal concentration was chosen as 0.5 µg based on the level of enrichment. If less than 0.5 µg of C19 was used, then the level of enrichment is lower than with 0.5 µg of C19, whereas using 1 µg did not improve the signal. These experiments were performed on sorted Sca1^{neg} cells from 10-week old virgin females.



Appendix 6. Accompaniment to Figure 4.7. Quality control analysis of ChIP-Seq data sets prior to analysis. (A) Bar graph show the total number of aligned reads for each sequencing library, and the proportion of distinct regions. (B) Bar graph shows the percentage of unique and not unique reads (unmapped or mapped to more than one genomic region) for each sequencing library. **Abbreviations:** P16.5, pregnant day 16.5 alveolar cells.



Appendix 7. Accompaniment to Figure 4.7. Quality control analysis of ChIP-Seq data sets prior to analysis. Chromosomal distribution of reads for each sequencing library. All libraries have equal coverage of all chromosomes (except Y and MT). **Abbreviations:** P16.5, pregnant day 16.5 alveolar cells; MT, mitochondrial genome.



Appendix 8. Accompaniment to Figure 4.10A. Complete functional annotation clustering table of C/EBPβ bound (Appendix 8A) and unbound (Appendix 8B) genes in Sca1^{pos} luminal cells generated using the DAVID bioinformatics resource with default parameters. Functional clusters were included in the figure if they had an enrichment score of ≥2.

Appendix 9. Accompaniment to Figure 4.10B&C. Complete functional annotation clustering table of C/EBPβ bound (Appendix 9A) and unbound (Appendix 9B) genes in Sca1^{neg} luminal cells generated using the DAVID bioinformatics resource with default parameters. Functional clusters were included in the figure if they had an enrichment score of ≥2.

Appendix 10. Accompaniment to Figure 4.11. Complete functional annotation clustering table of C/EBPβ bound (Appendix 10A) and unbound (Appendix 10B) genes in P16.5 alveolar cells generated using the DAVID bioinformatics resource with default parameters. Functional clusters were included in the figure if they had an enrichment score of ≥2.